Curso de Cálculo para ¿ingenieros?

Pedro Fortuny Ayuso

Curso 2011/12, EPIG, Gijón. Universidad de Oviedo
E-mail address: fortunypedro@uniovi.es

3

CAPÍTULO 1

Introducción

1. Comentarios superficiales

En este curso vamos a aprender esencialmente dos cosas:

- Acotar,
- Optimizar.

Están relacionadas, pues optimizar consiste en encontrar "la cota máxima o mínima", la mejor de todas en cierto sentido.

¿Pero acotar qué?

Todo lo relacionado con funciones de una variable real (para lo cual haremos uso de las nociones de continuidad, derivabilidad e integración), sucesiones y funciones de varias variables reales (en este caso solo estudiaremos lo relacionado con la "derivabilidad", que se denomina *diferenciación*).

En la vida real los problemas físicos e ingenieriles se pueden dividir en tres tipos:

1. Los que responden a una pregunta directa (¿cuánto cuesta...?).
2. Los que consisten en buscar un valor óptimo (¿cuál es la máxima temperatura que alcanza cierto cuerpo sujeto a ciertas condiciones? ¿cuál es el volumen máximo que se puede encerrar con una plancha de un metro cuadrado?...)
3. La búsqueda de cotas. Quizás estos son los más realistas en muchos casos. Ejemplos: ¿tengo suficiente agua para alimentar a todos estos camellos? ¿resistirá este andamio el peso de todos estos obreros? ¿será este cable suficientemente ancho para aguantar la potencia eléctrica que pasará por él?

Todos los ejemplos expuestos pueden enunciarse de manera sintética haciendo uso del "lenguaje" de las funciones, la continuidad, la integración, la derivación, etc... Por eso es interesante esta asignatura, porque os provee del lenguaje adecuado para hacer las preguntas de la manera correcta (es el *ingeniero* el que debe trasladar el problema "real" al lenguaje matemático: los matemáticos *no sabemos nada de la realidad*).

Acotar, aunque lo veremos con detalle en los múltiples ejercicios que haremos, es una de las tareas más importantes de un ingeniero, porque habitualmente no hay manera de conocer la solución exacta de un problema (puede que incluso ni siquiera tenga sentido buscarla) pero sí hay forma de saber *entre qué dos valores* puede estar dicha solución,

o bien (lo cual es suficiente en ciertos casos), más de cuánto o menos de cuánto es ese valor (dar una cota inferior o superior, respectivamente).

Eso por un lado.

2. Conocer las herramientas

Por otro lado, sería tonto no utilizar las herramientas de que disponemos hoy en día. En el navegador de un *smartphone*, o en un ordenador, váyase a `www.wolframalpha.com` (esta dirección es de conocimiento obligatorio en esta asignatura, la preguntaré en el examen). Es la ventana de un buscador, pero no es un buscador. Es mucho más.

Escríbase en el área de texto lo siguiente (el símbolo entre la `x` y el 2 es un acento circunflejo:

```
int(x/(x^2+1),x)
```

y dése al botón "=" (o Enter). Mágicamente, se obtiene el resultado de la fórmula

$$\int \frac{x}{x^2+1}\,dx = \frac{1}{2}\log(x^2+1) + c$$

amén de un montón de información añadida que puede ser utilísima.

Otro ejemplo:

```
limit(n log(n^4-1)/log(n!),n,inf)
```

calcula el límite de la sucesión entre paréntesis, que para vosotros puede ser sencillo, siempre y cuando conozcáis la fórmula de Stirling.

En las prácticas de ordenador utilizaréis uno de los programas para cálculo simbólico que hay en el mercado. Posiblemente Maxima o Matlab. Pero para cuentas "sobre la marcha" sin demasiadas complicaciones, WolframAlpha es una herramienta imprescindible. Además es mucho más que una calculadora científica. Varios ejemplos:

```
isoperimetric problem
```

```
max(-x^2 - y^2 +33)
```

```
jupiter
```

```
the lord of the rings
```

Claro que hay que manejarlo en inglés, pero hoy día vivimos en ese idioma.

CAPÍTULO 2

Funciones de una variable real

En este capítulo estudiamos (repasamos) los conceptos fundamentales del cálculo diferencial de una variable real. El objetivo es ser capaz de hacer estudios *cualitativos* locales de las funciones reales de una variable real, y aprender a utilizar los teoremas fundamentales para acotar.

Sin duda, los resultados más importantes son:

- El Teorema de Bolzano (y sus corolarios).
- El Teorema del valor medio de Lagrange (y sus corolarios de acotación).
- La noción de punto crítico y su discernimiento.
- El Polinomio de Taylor y las aproximaciones y cotas del resto.

Todos ellos *habrán de saberse de memoria*, porque se utilizan constantemente en problemas *de la vida real*. Quizás alguno se presente "completamente" demostrado, pero lo que es esencial es conocer su enunciado *con precisión*.

Antes de definir la noción de función, etc. repasamos someramente la noción de número real —esencialmente repasamos el *Axioma de los intervalos encajados*—, poniendo énfasis en el principio del continuo.

1. Qué es $\mathbb{R}$

El conjunto básico sobre el que vamos a trabajar en toda esta asignatura —es decir, los *números* que vamos a utilizar— es $\mathbb{R}$, el *cuerpo de los números reales*.

Se puede definir $\mathbb{R}$ de la siguiente manera: es el conjunto más pequeño que cumple las siguientes condiciones:

1. Los racionales $\mathbb{Q}$ están todos en $\mathbb{R}$ (esto se dice $\mathbb{Q} \subset \mathbb{R}$, están contenidos en él).
2. En $\mathbb{R}$ se puede sumar $+$, restar $-$, multiplicar $\cdot$ y dividir $/$ (salvo por 0) y el 0 y el 1 cumplen sus funciones de elementos neutros, etc...
3. El conjunto $\mathbb{R}$ está totalmente ordenado (dados dos números $x, y \in \mathbb{R}$, se tiene que o bien $x - y > 0$ o bien $y - x > 0$ o bien $x = y$), y se cumplen las condiciones normales del orden (luego las enunciamos).
4. **Axioma de los intervalos encajados**: (esto es lo que diferencia a $\mathbb{R}$ de $\mathbb{Q}$): si a_n y b_n son dos sucesiones de números reales

tales que $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$ (es decir, a_n es creciente, b_n es decreciente y las a_n son todas menores que las b_n), y $b_n - a_n$ tiende a cero[1], entonces

$$\bigcap [a_n, b_n] = \{p\}$$

para cierto $p \in \mathbb{R}$.

Gráficamente, el axioma de los intervalos encajados dice que una sucesión de intervalos *cerrados* y *cada vez más estrechos* contenidos cada uno en el anterior contiene un número real (y solo uno) en su interior. Esto es una manera "precisa" de decir que $\mathbb{R}$ está *lleno*: no solo que entre dos números reales hay otro número real (pues esto pasa para los racionales, también), sino que si uno va recortando segmentos cada vez más estrechos hasta anchura cero, uno siempre *encuentra algo ahí*.

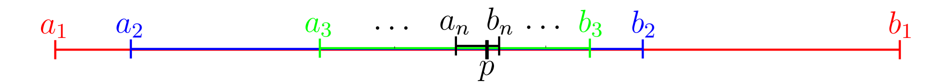


FIGURA 1. Imagen del axioma de los intervalos encajados

Es bien conocido que el axioma de los intervalos encajados no es cierto en $\mathbb{Q}$ (es conocido, más o menos y en otro contexto, desde Pitágoras, aunque no lo dijera así). Si l es la longitud de la diagonal de un cuadrado de lado 1, se pueden ir dibujando segmentos de longitud racional más pequeñas que l, pero que se vayan acercando a ese valor y otros de longitud racional máyor, pero *no hay ninguno de longitud racional* l, porque $\sqrt{2}$ no es un número racional. ¿Cómo se demuestra esto? Es bien sencillo, si se sabe algo sobre factorizar números enteros y cómo se calculan cuadrados de números racionales.

Recordamos, antes de seguir, una noción que ha de ser conocida ya:

DEFINICIÓN 1. Un *intervalo abierto* es el conjunto de números mayores que uno a y menores que otro b, *estrictamente*. Se denota (a, b). Se llama intervalo cerrado $[a, b]$ al intervalo abierto junto con los extremos a y b (es decir, tiene dos elementos más que el abierto, salvo en el caso $a = b$, claro).

Hablaremos de *intervalo extendido* cuando admitamos que uno de los extremos sea $+\infty$ o $-\infty$ (o ambos), pero por lo general cualquier intervalo sin más tendrá los extremos en $\mathbb{R}$. Si uno de los extremos tiene un corchete, se quiere decir que se incluye ese punto en el intervalo: $[a, b) = (a, b) \cup \{a\}$, etc.

Del principio de los intervalos encajados se concluyen otros equivalentes (cualquiera de estos se puede utilizar como axioma básico para construir $\mathbb{R}$ a partir de $\mathbb{Q}$):

[1]Esta noción no la hemos definido, pero asumimos que el lector está familiarizado con ella.

- Cualquier subconjunto $K \subset \mathbb{R}$ acotado superiormente admite un supremo. Es decir, si hay un número M tal que $M \geq k$ para cualquier $k \in K$, entonces existe un M_0 con la misma propiedad que es el más pequeño de todos. (Principio del supremo).
- Cualquier subconjunto $K \subset \mathbb{R}$ acotado inferiormente admite un ínfimo.
- Si $A \cup B = \mathbb{R}$ y ambos son no vacíos y cualquier elemento de a es menor que cualquier elemento de B, entonces existe $t \in \mathbb{R}$ tal que $a \leq t$ para todo $a \in A$ y $t \leq b$ para todo $b \in B$ (particiones de Dedekind).

En esos enunciados se ha utilizado la noción de "cota", "supremo" e "ínfimo". Definimos:

Definición 2. Dado un conjunto $C \subset \mathbb{R}$, se dice que C está *acotado superiormente* si existe un $M \in \mathbb{R}$ tal que $M \geq c$ para todo $c \in C$. Se dice *acotado inferiormente* si ocurre lo mismo pero con $M \leq c$ para todo $c \in C$. Se dice *acotado* si está acotado inferior y superiormente.

Cualquier número M que cumple la condición de la definición se llama *cota superior* (o *inferior*) de C. Un conjunto acotado superiormente "no llega hasta infinito", mientras que uno acotado inferiormente "no llega hasta menos infinito". Un conjunto acotado, es uno que se "cabe entre dos números reales".

Definición 3. El *supremo* de un conjunto $K \subset \mathbb{R}$ es, *si existe*, la menor de las cotas superiores. El *ínfimo* es la mayor de las cotas inferiores. Si el supremo está en K, entonces se llama *máximo* y si el ínfimo está en K, entonces se llama *mínimo*.

Conjuntos con supremo pero no máximo: cualquier intervalo abierto (a, b).

Algunos ejemplos más:

- El conjunto de los números de la forma $\frac{1}{n}$, cuando $n \in \mathbb{N}$. Está acotado superiormente, el supremo es 1, que es máximo. Está acotado inferiormente, el ínfimo es 0, pero *no es mínimo.*

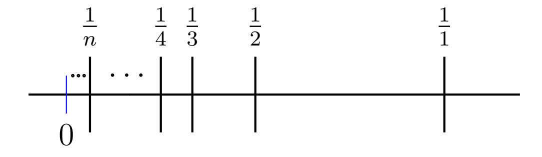


Figura 2. El conjunto $\{1/n\}_{n=1}^{\infty}$ y su ínfimo (0), que no es mínimo.

- El conjunto $\{x \in \mathbb{R} | x^2 < 2\}$ está acotado superiormente, *no inferiormente*, 3 es una cota superior y $\sqrt{2}$ es el supremo.

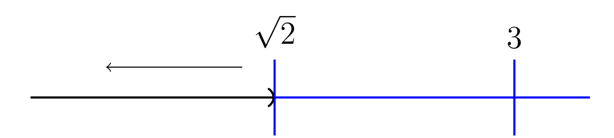


FIGURA 3. El conjunto $\{x|x^2 < 2\}$ (en negro), el supremo ($\sqrt{2}$) y una cota (3).

2. Función, función inversa

El objeto de estudio del Cálculo son las *funciones*. La definición precisa es poco inteligible:

DEFINICIÓN 4. Una *función* real de variable real es un subconjunto $K \subset \mathbb{R} \times \mathbb{R}$ tal que si $(x, y_1), (x, y_2) \in K$ entonces $y_1 = y_2$.

En castellano: una *asignación* de un número y a cada elemento de un subconjunto $X \subset \mathbb{R}$. A cada elemento $x \in X$, se le asigna $y = f(x)$, si la función se llama f. Se denota así:

$$\begin{array}{rcl} f : X & \longrightarrow & \mathbb{R} \\ x & \longmapsto & f(x) \end{array}$$

Por ejemplo, la función que a cada elemento de $\mathbb{R}$ le asigna su cuadrado más 5. En este caso, el "origen" es $\mathbb{R}$ y el conjunto final es $\mathbb{R}$:

$$\begin{array}{rcl} f : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^2 + 5 \end{array}$$

La función que a cada real positivo le asigna su raíz cuadrada positiva está definida solo en la parte positiva de $\mathbb{R}$ (junto con el 0), que designamos $\mathbb{R}_{\geq 0}$:

$$\begin{array}{rcl} f : \mathbb{R}_{\geq 0} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \sqrt{x} \end{array}$$

(se suele poner así, sin signo antes de la raíz cuadrada).

DEFINICIÓN 5. El *dominio* de definición de una función f es el conjunto en el que está definida. El *recorrido* es el subconjunto de $\mathbb{R}$ "de valores tomados por la función", es decir, de aquellos $y \in \mathbb{R}$ para los que existe un x en el dominio con $f(x) = y$.

Las siguientes nociones han de ser conocidas:

DEFINICIÓN 6. Una función $f : X \to Y$ es

Inyectiva: si $f(x_1) = f(x_2)$ implica $x_1 = x_2$. Es decir, si cada valor es tomado solo una vez.

Sobreyectiva: si para cada $y \in Y$ hay un x tal que $f(x) = y$. Es decir, si "llena" el conjunto Y.

Biyectiva: Si es inyectiva y sobrellectiva (llena el conjunto Y y además, no se "repiten" valores).

Dadas dos funciones $f : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R} \to \mathbb{R}$, se llama *composición* de f con g a la función $f \circ g$ dada por $f \circ g(x) = f(g(x))$.

Una función biyectiva tiene *inversa*:

DEFINICIÓN 7. Si $f : X \to Y$ es una función biyectiva, se llama *inversa de* f, y se denota $f^{-1} : Y \to X$ a la función que asocia a cada $y \in Y$ el único elemento $x \in X$ tal que $f(x) = y$.

Gráficamente, la función inversa de una función real de variable real se dibuja "reflejando la gráfica en la recta $y = x$":

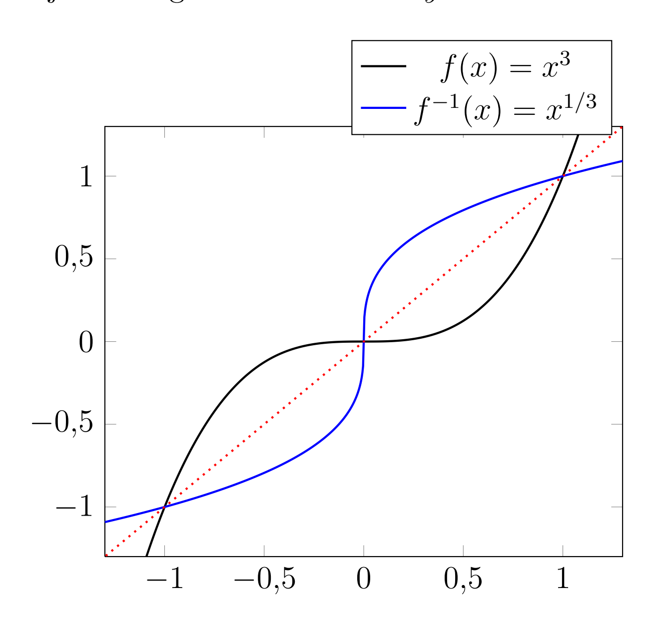


FIGURA 4. Función inversa (en azul) de la función x^3 (en negro). Ambas son, obviamente, biyectivas.

3. Límite de una función

3.1. Infinitésimo. Todo nuestro desarrollo de la continuidad, derivabilidad y (en el futuro) diferenciabilidad de funciones se basará en la noción de "función que se hace muy pequeña" cuando la variable es pequeña. Llamaremos a esto un "infinitésimo" o una "cantidad diferencial" (pero este término no lo usaermos mucho):

DEFINICIÓN 8. Un *infinitésimo en* 0 o una *cantidad diferencial* (o, brevemente, una *diferencial*) es una aplicación $\mathrm{d} : (-a, a) \to \mathbb{R}$ de un entorno de 0 en $\mathbb{R}$ tal que para cualquier número (por pequeño que sea) $\varepsilon > 0$, existe un entorno $(-c, c)$ tal que $|\mathrm{d}(h)| < \varepsilon$ en todo $(-c, c)$.

El "en 0" se verá en el futuro a qué se debe. De momento, no lo utilizaremos nunca porque hasta que los necesitemos, no hablaremos de infinitésimos en otros puntos.

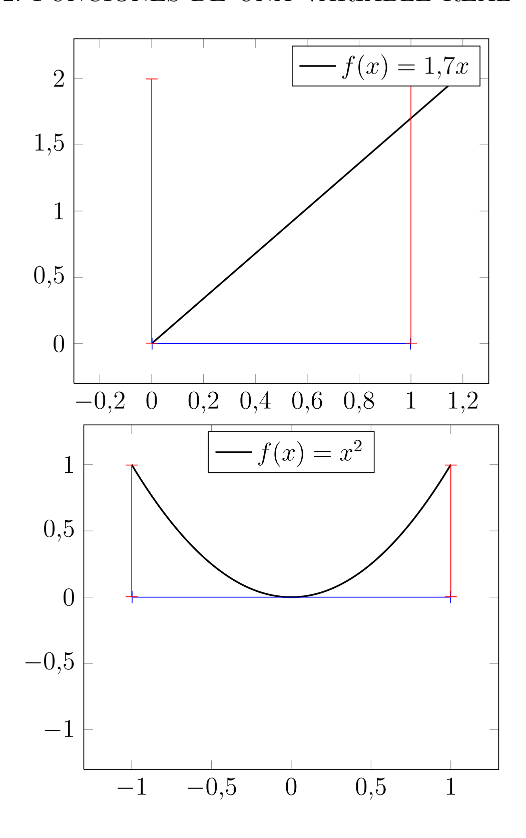


FIGURA 5. La función f es inyectiva de $[0, 1]$ a $[0, 2]$ pero no sobre. La función g es sobreyectiva, pero no inyectiva de $[-1, 1]$ a $[0, 1]$.

Es decir, un *infinitésimo* es una función que "es más pequeña" (en valor absoluto) que cualquier cantidad positiva. Es "más pequeña", por así decir, "a partir de un momento".

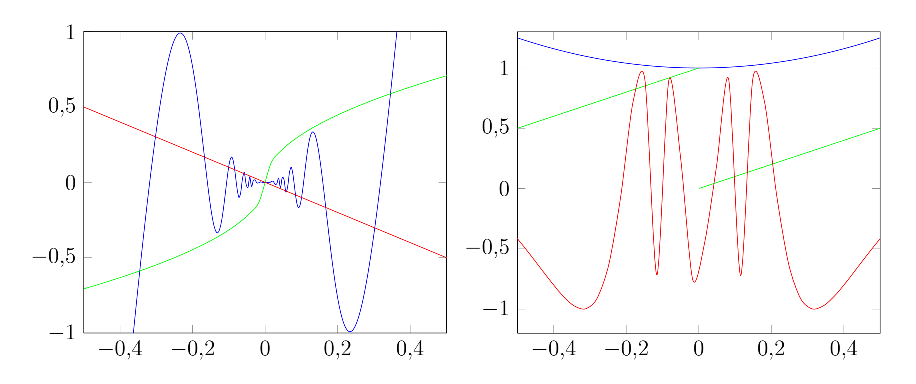


FIGURA 6. Infinitésimos y funciones que no lo son.

La utilidad de esta definición (que **cae en el examen**) es que todas las nociones *dinámicas* se pueden enunciar utilizando infinitésimos, en lugar de utilizar permanentemente el lenguaje de épsilons y deltas, que es mucho más farragoso. Solo hay que entender, de una vez por todas,

que una cantidad diferencial representa una función que "se acerca a cero según la variable se acerca a cero".

Vamos a utilizar muchas veces las siguientes propiedades:

TEOREMA 1 (Propiedades de los infinitésimos). *Sean* $f, g : (-a, a) \to \mathbb{R}$ *dos infinitésimos. Entonces:*

- *La suma y la resta* $f \pm g$ *son infinitésimos.*
- *El producto* fg *es un infinitésimo.*
- *La composición* $g \circ f$ *es un infinitésimo*[2]*.*
- *Si* $f(x) \neq 0$ *entonces el inverso* $1/f$ *es un* infinito*: dado cualquier número* $M > 0$ *existe un intervalo* $(-c, c)$ *tal que* $|f(x)| > M$ *para todo* $x \in (-c, c)$ *(es decir, el inverso de un infinitésimo se hace infinitamente grande".)*

La noción clave de toda la asignatura es la de límite de una función en un punto. Antes precisamos una noción

DEFINICIÓN 9. Se dice que a es un punto *adherente* a un conjunto X si cualquier intervalo abierto que contiene a a corta a X en algún punto (es decir, si a está en X o *muy cerca de él*). Se escribe $a \in \bar{X}$. Se dice que a es un punto de acumulación si cualquier intervalo que contiene a a corta a X *en un punto distinto de* a. Se escribe $a \in X'$.

DEFINICIÓN 10. Sea $f : X \to \mathbb{R}$ una función real de variable real (es decir, $X \subset \mathbb{R}$) y sea $a \in X'$ un punto de acumulación de X. Se dice que l es el *límite de* f *cuando* x *tiende a* a y se escribe

$$\lim_{x \to a} f(x) = l$$

si existe un infinitésimo d tal que $f(a + h) = l + \mathrm{d}(h)$ para $|h| > 0$ (y siempre que $a + h \in X$, obviamente).

Es decir, l es el límite de f en a si $f(a + h)$ es muy cercano a l cuando h es muy próximo a cero. Si esto ocurre solo por "un lado", se habla de *límite por la derecha* y de *límite por la izquierda*:

DEFINICIÓN 11. Si existe un infinitésimo d tal que $f(a + h) = l + \mathrm{d}(h)$ para $h > 0$ entonces se dice que f tiene límite l *por la derecha*. Si ocurre lo mismo pero para $h < 0$, entonces se dice que el límite es por la izquierda. Obviamente, si ocurren ambas cosas es porque el límite de f en a es l.

NOTA 1. La definición que hemos dado es equivalente a la clásica definición *para todo* $\varepsilon > 0$ *existe un* $\delta > 0$ *tal que* $|h| < \delta$ *implica* $|f(a + h) - l| < \varepsilon$, pero *pienso* que es más sencilla de entender, por ser más natural (un número es el límite si la función vale ese número más algo muy pequeño cerca del punto). Se puede imaginar la noción "clásica" de límite como una condición "dialogada". Supongamos que

[2] Donde está definida, que es en un entorno un poco más pequeño

hay dos "jugadores", Ana y Basilio. El número l es el límite de f en a si para cualquier elección que haga Ana de una distancia en "vertical", Basilio puede encontrar una distancia en "horizontal" tal que cualquier punto que esté más cerca de a de lo que dice Basilio, tiene la imagen más cerca de l de lo que dice Ana.

La función $\operatorname{sen}(1/x)$ para $x > 0$ no tiene límite en 0: dado, por ejemplo, $\varepsilon = 1/2$, hay puntos x dentro de cualquier entorno $(0, \delta)$ para los cuales $|\operatorname{sen}(1/x) - 0| > 1/2$ (por ejemplo, todos los puntos de la forma $\frac{2}{(2k+1)\pi}$, para $k \in \mathbb{N}$).

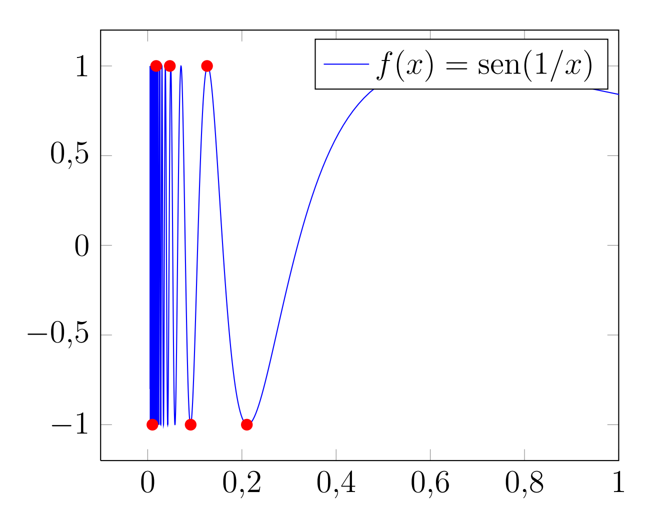


FIGURA 7. La función $f(x) = \operatorname{sen}(1/x)$ no tiene límite en 0.

Existen también las nociones de *límite infinito* y *límite en el infinito*. Antes de nada, utilizamos parte del Teorema 1 para definir la noción de *función que se hace infinito*.

DEFINICIÓN 12. Se dice que $f : (a - \varepsilon, a + \varepsilon) \to \mathbb{R}$ es un *infinito en a* si dado cualquier $M > 0$, hay un δ tal que $f(x) > M$ para $x \in (a - \delta, a + \delta)$ y $x \neq a$. Si solo ocurre para $x \in (a, a + \delta)$, se habla de *infinito por la derecha de a*, y si es para $(a - \delta, a)$, de *infinito por la izquierda de a*. Si lo que ocurre es que $f(x) < -M$, entonces se habla de un *menos infinito* (pero nunca diremos esto).

DEFINICIÓN 13. Se dice que $f : X \to \mathbb{R}$ tiene límite $+\infty$ en un punto c de acumulación de X si existe un infinito *positivo* $I : (-h, h) \to \mathbb{R}$ tal que

$$f(c + h) = I(h)$$

para $h \in (-a, a)$ (y, como siempre, $c + h \in X$, pero esto es obvio). Si el infinito es *negativo* ($I(x) < 0$) entonces se habla de límite $-\infty$.

Para definir el límite en infinito, recurrimos en parte a la noción clásica, pero nunca más lo haremos:

Definición 14. Se dice que $f : X \rightarrow \mathbb{R}$ tiene límite l en $+\infty$ si para cualquier $\varepsilon > 0$ existe $K > 0$ tal que $|f(x) - l| < \varepsilon$ para todo $x > K$.

Se dice que el límite en $+\infty$ es $+\infty$ para todo $M > 0$ existe $K > 0$ tal que $f(x) > M$ si $x > K$.

Para $-\infty$ y todo lo demás, no hay más que cambiar los signos...

Al final todo termina en una *mecánica* que el alumno ya debería conocer.

3.2. Mecánica del cálculo de límites. No vamos a estudiarla en detalle en estos apuntes. Este es uno de los pocos apartados en que vamos a dar por supuesto que el alumno ya conoce.

4. Continuidad, límites

La última definición de la sección anterior, la de *límite* es el primer paso para poder hacer un análisis local de las funciones, es decir, para "distinguir" qué funciones tienen un comportamiento "normal" y cuáles no. El objetivo de una definición es precisamente ese: dividir *para aclarar*. ¿Qué aclara la noción de límite? La idea (y esta es la única idea importante del análisis) de que *cuando uno se aleja muy poco del punto a, la función se aleja muy poco de l*, eso es lo que significa la expresión

$$\lim_{x \to a} f(x) = l.$$

La clave es precisamente esa: cuando uno está *muy cerca de a*, la función está *muy cerca de l*. Esto es distinto de decir que f *se acerca "todo lo que yo quiera" a l cuando x se acerca a a*, porque esta frase es verdad para la función $\operatorname{sen}(1/x)$ (ver figura 7) respecto del valor 1 cuando x se acerca a 0, pero es obvio que esta función *no tiene límite* 1 en 0. De hecho, como se ve en su gráfica, *no puede tener ningún límite*: cuando x se aleja un poco de 0, el valor $f(x)$ puede acercarse a cualquier valor entre -1 y 1.

Con esta idea, es evidente que la siguiente función tiene límite en $a = 1$, y el límite es 2, aunque la función en 1 valga 0,83. Esto es un ejemplo del principio primero de los límites: **El valor de la función en el punto no tiene nada que ver con el límite**.

Precisamente la idea de que *el límite debe estar relacionado con el valor de la función* es lo que lleva a la definición de *continuidad* de una función en un punto. Tanto en la Figura 8 como en la 7, se observa que el comportamiento de la gráfica cerca del punto en que se está mirando el límite es "extraño". Esa extrañeza se llama "discontinuidad".

Definición 15. Se dice que una función $f : X \rightarrow \mathbb{R}$ es continua en $a \in X$ si f tiene límite en a y además

$$\lim_{x \to a} f(x) = f(a)$$

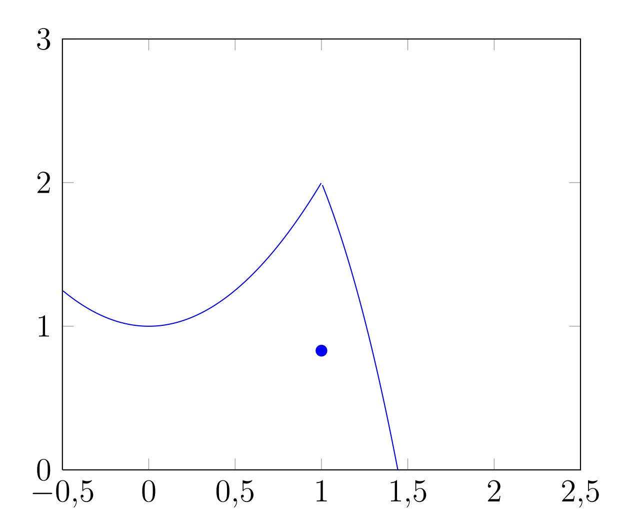


FIGURA 8. Función con límite en $a = 1$, pero con valor distinto del límite.

es decir, si el límite coincide con el valor de la función en el punto.

Si f es continua en todos los puntos de X, entonces se dice que f es *continua en* X.

Uniendo esta definición y la de límite, es mucho más sencillo enunciarlo así:

DEFINICIÓN 16. Se dice que $f : X \to \mathbb{R}$ es conínua en $a \in X$ si existe un infinitésimo d tal que

$$f(a + h) = f(a) + \mathrm{d}(h)$$

para $h \in (-c, c)$ y $a + h \in X$.

Si solo ocurre para $h \in (0, c)$, se habla de *continuidad por la derecha* y si es para $h \in (-c, 0)$, de *continuidad por la izquierda*.

Cuando X es como un intervalo, la continuidad *en todo* X se puede resumir gráficamente diciendo que la gráfica f *se puede pintar sin levantar el lápiz del papel*.

También se consideran los límites *por la derecha y por la izquierda* de un punto $a \in \mathbb{R}$ y se habla de continuidad por la derecha y por la izquierda. No vamos a insistir: una función es continua en un punto si y solo si los límites laterales coinciden con el valor de la función en el punto.

4.1. Las funciones racionales son continuas. Es muy sencillo comprobar lo siguiente:

- Las funciones constantes $f(x) = c$ para $c \in \mathbb{R}$ son continuas en todo $\mathbb{R}$. Su límite en cada punto es c.
- La función $f(x) = x$ es continua en todo $\mathbb{R}$. Su límite en cada punto es x. Sería oportuno tratar de probar esto.

- Si $f(x)$ y $g(x)$ son continuas en a, entonces $f(x) + g(x)$ y $f(x)g(x)$ son continuas en a y sus límites son, en cada punto a, $f(a) + g(a)$ y $f(a)g(a)$.
- Si $f(x)$ es continua en a y $\lim_{x\to a}(f(x)) \neq 0$, entonces $1/f(x)$ es continua en a y su límite en a es $1/f(a)$. Probar esto puede ser una buena idea.
- Si $r > 0$ es un número real positivo, entonces r^x es continua en todo $\mathbb{R}$ y su límite es r^x en cada punto. Esto es sencillo de comprobar.
- Si $f(x)$ es continua en $[c, d]$ y $g(x)$ es continua en $[f(c), f(d)]$ (bien ordenado) entonces $g(f(x))$ es continua en $[c, d]$ y el límite en cada punto $a \in [c, d]$ es $g(f(a))$. Este resultado es bien fácil de "demostrar" utilizando infinitésimos. Solo se ha de comprobar que $g(f(a + h))$ es $g(f(a)) + \mathrm{d}(h)$, donde d es cierto infinitésimo, pero como f es continua, hay un infinitésimo d_1 tal que:
$$g(f(a + h)) = g(f(a) + \mathrm{d}_1(h))$$
y como g es continua en $f(a)$, hay otro d_2 tal que
$$g(f(a + \mathrm{d}_1(h))) = g(f(a) + \mathrm{d}_1(h)) =$$
$$g(f(a)) + \mathrm{d}_2(\mathrm{d}_1(h)) = g(f(a)) + \mathrm{d}(h)$$
donde d es *infinitamente pequeño*[3], así que $g \circ f$ es continua. Este razonamiento *es correcto* pero el lenguaje no lo conocéis. No importa, así se hacen las cosas.
- Etcétera (no voy a seguir con todas estas nociones bien conocidas).

A partir de la lista de resultados anteriores se deduce, por ejemplo, que todas las funciones racionales
$$f(x) = P(x)/Q(x)$$
son continuas en todo $\mathbb{R}$ salvo quizás en los ceros de $Q(x)$ (de hecho, si la fracción racional está simplificada, en los ceros de $Q(x)$ hay siempre una discontinuidad, pero solo si f está simplificada).

Pero también un montón de otras funciones lo son. Ahora, la pregunta es ¿cómo son las funciones discontinuas? Hay tres tipos: dos ya se han visto en las Figuras 7 y 8: son dos funciones que no pueden dibujarse sin levantar el lápiz del papel. Hay otro tipo: cuando la función *se va a infinito*:

DEFINICIÓN 17. Una función discontinua en a se dice que tiene una discontinuidad

Evitable: Si la función tiene límite en a pero $f(a)$ no es igual a dicho límite.

[3]Porque *la composición de infinitésimos es un infinitésimo*, como se vio en el Teorema 1.

De salto finito: Si el límite por la derecha es distinto del límite por la izquierda y ambos existen.

De salto infinito: Si los límites laterales existen y uno de ellos es infinito.

Esencial: Si uno de los dos límites laterales ni existe ni es infinito.

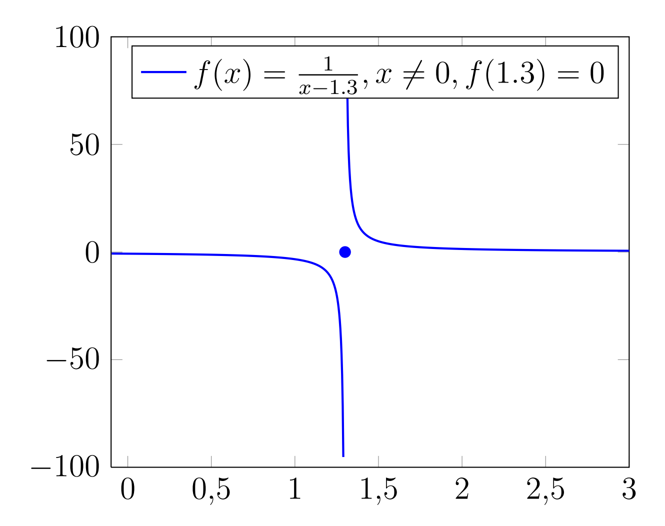


FIGURA 9. Función con límite $+\infty$ por la derecha, $-\infty$ por la izquierda. Obviamente, discontinua en 1.3.

5. Los teoremas básicos de la continuidad

Hay tres resultados *que hay que saberse de memoria, tal y como están enunciados,* relativos a propiedades *obvias* de las funciones continuas en intervalos cerrados, que son de la máxima *utilidad*. Previamente se demuestra siempre una propiedad clave: una función continua que no es nula en un punto c tiene signo constante en un entorno de c:

LEMA 1. *Si $f : (a, b)$ es continua en un punto $c \in (a, b)$ y $f(c) \neq 0$, entonces existe $\delta > 0$ tal que el signo de f en $(c - \delta, c + \delta)$ es el mismo que en c.*

DEMOSTRACIÓN. Vamos a demostrarlo, aunque sea para tener una idea. En el punto c, la función vale $f(c)$ que, por ejemplo, es mayor que cero (si es menor, se hace un razonamiento análogo). Como f es continua en c, se sabe que existe un infinitésimo d tal que

$$f(c + h) = f(c) + \mathrm{d}(h).$$

Pero por definición de infinitésimo, hay un entorno $(-\delta, \delta)$ tal que $|\mathrm{d}(h)| < f(c)$ en todo ese entorno, de manera que

$$f(c) + \mathrm{d}(h) > 0$$

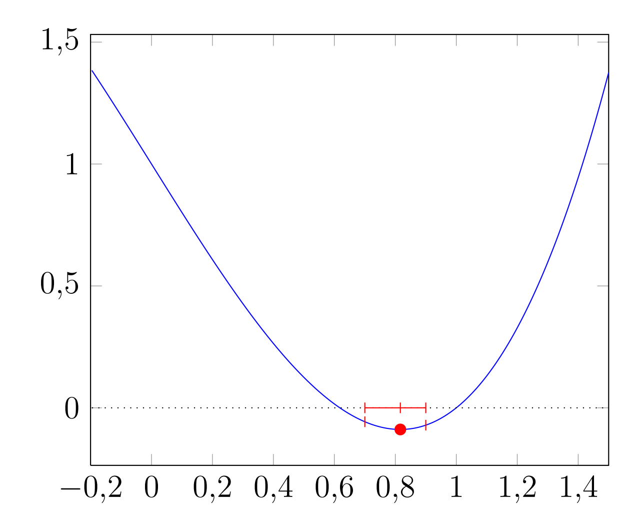


FIGURA 10. El signo es constante en un entorno de un punto de continuidad.

para $h \in (-\delta, \delta)$ y por tanto

$$f(c+h) = f(c) + \mathrm{d}(h) > 0$$

para los mismos h. Así que si $x \in (c-\delta, c+\delta)$, se cumple que $f(x) > 0$. □

TEOREMA 2 (Teorema de Bolzano). *Si $f : [a,b] \to \mathbb{R}$ es una función continua definida en un intervalo cerrado y $f(a)$ y $f(b)$ tienen distinto signo (es decir, $f(a) \cdot f(b) < 0$), entonces existe* al menos *un punto $x_0 \in (a,b)$ tal que $f(x_0) = 0$).*

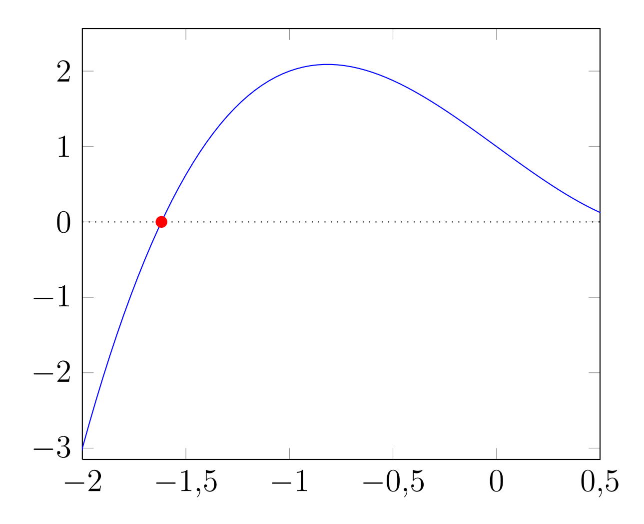


FIGURA 11. Teorema de Bolzano.

TEOREMA 3 (Teorema del valor intermedio). *Sea* $f : [a, b] \to \mathbb{R}$ *una función continua en* $[a, b]$. *Si* c *es un valor entre* $f(a)$ *y* $f(b)$, *entonces existe* $x_0 \in [a, b]$ *tal que* $f(x_0) = c$.

Téngase en cuenta que en este resultado, el valor c puede ser tanto $f(a)$ como $f(b)$ y por ello x_0 puede ser uno de los extremos. La prueba de este teorema es una tontería, utilizando la función $g(x) = f(x) - c$ y el Teorema de Bolzano.

El siguiente resultado significa que "una función que se puede dibujar sin levantar el lápiz tiene cabe en un papel pequeño":

LEMA 2. *Si* $f : [a, b] \to \mathbb{R}$ *es continua, entonces el conjunto* $f([a, b])$ *de imágenes está acotado superior e inferiormente.*

TEOREMA 4 (Teorema de Weierstrass). *Sea* $f : [a, b] \to \mathbb{R}$ *una función continua en* $[a, b]$. *Sean* $M = \text{máx}\{f(x)\big| x \in [a, b]\}$ *y* $m = \text{mín}\{f(x)\big| x \in [a, b]\}$ *el máximo y el mínimo de los valores de* f *en* $[a, b]$. *Entonces existen puntos* $c, C \in [a, b]$ *tales que* $f(c) = m$ *y* $f(C) = M$.

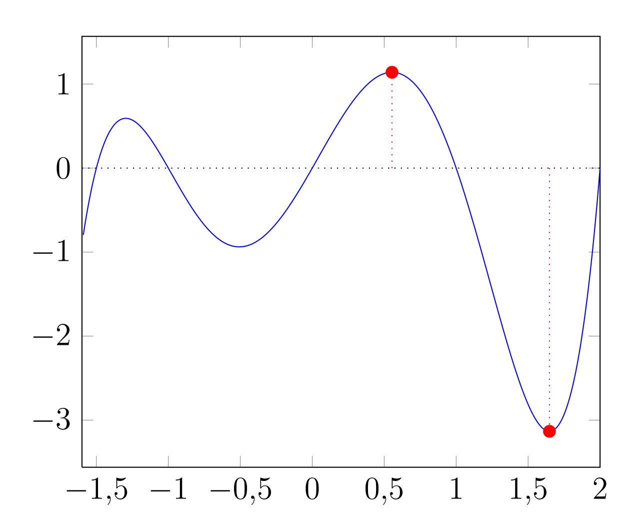


FIGURA 12. Teorema de Weierstrass.

Un corolario clásico del Teorema de Bolzano es el siguiente enunciado: si subes una montaña un día, comenzando a una hora determinada y al día siguiente la bajas comenzando a la misma hora y terminando a la misma, entonces hay un momento de la bajada en que estabas justamente en ese sitio al subir:

LEMA 3. *Si* $f : [a, b] \to \mathbb{R}$ *y* $g : [a, b] \to \mathbb{R}$ *son tales que* $f(a) < f(b)$, $g(a) = f(b)$ *y* $g(b) = f(a)$ *entonces existe un punto* $c \in (a, b)$ *tal que* $g(c) = f(c)$.

Demostrar esto es muy fácil: en vez de hacer la subida un día y la bajada al siguiente, hace una persona la subida y otra la bajada: en un momento dado han de cruzarse (porque se supone que uno va de

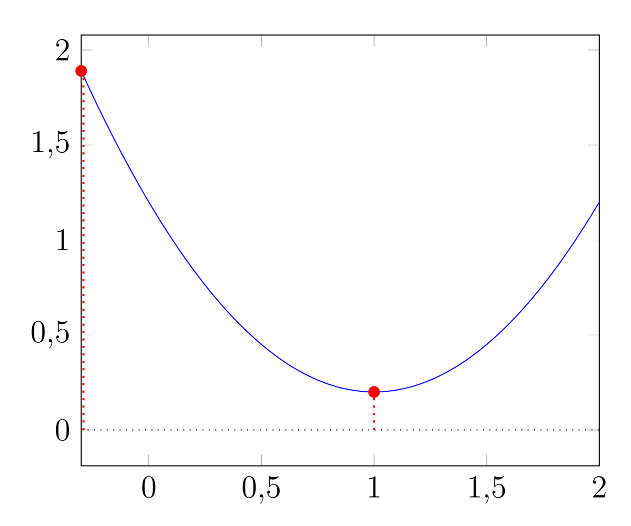


FIGURA 13. Teorema de Weierstrass: el máximo o mínimo puede estar en uno de los extremos.

manera *continua* por una montaña, no se *teleporta*, al menos a día de hoy (2011)).

6. Derivabilidad

La noción de *continuidad* es la primera noción *gráfica* del Cálculo: discierne las funciones que “no tienen saltos”. O, visto desde otra perspectiva, es la primera noción *infinitesimal*: discierne las funciones que “crecen muy poco” cuando la variable “crece poco” (recordemos, $f(a+h) = f(a) + \mathrm{d}(h)$). Pero es obvio que no es la única noción local que se puede estudiar. Desde el punto de vista físico (y la Física es quien mueve a las Matemáticas salvo en excepciones muy puntuales) es conveniente saber cómo identificar funciones que “se comportan bien” dinámicamente. La primera noción dinámica más allá de la continuidad es, claramente, “dar las curvas sin violencia”. Es decir, *no tener esquinas.*

Una “esquina” es un lugar en que “por un lado se va en una dirección y por otro lado se va en otra”.

Dirección = recta tangente

Una recta “no vertical” (veremos qué pasa con las verticales) en el plano tiene ecuación

$$y = ax + b$$

donde a es “la pendiente” (a, de hecho, es justamente la pendiente en el sentido de las carreteras: una cuesta del 8 % se escribe como una recta $y = 0{,}08x + \cdots$) y b se puede calcular una vez que se conoce un punto por el que pasa la recta y su pendiente. Es obvio que *la dirección* de una recta es el parámetro a (las rectas paralelas tienen todas la misma dirección), así que para hablar de *dirección* solo nos interesa

conocer *la pendiente* de la recta tangente a la gráfica de una función. Intuitivamente es fácil convencerse de que la recta tangente a un punto es *el límite* (si existe) de las secantes que pasan por ese punto y otro *infinitamente próximo* a él.

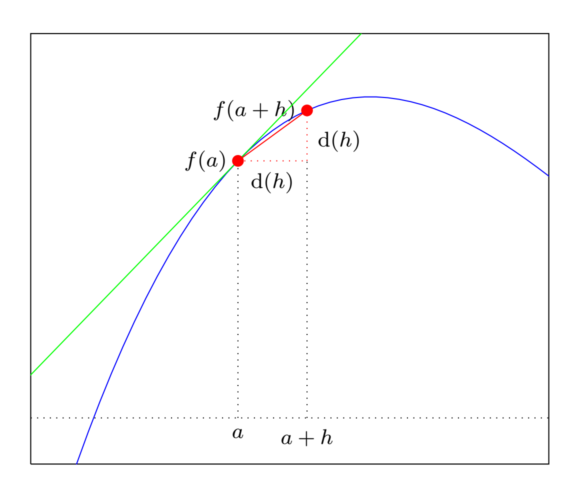


FIGURA 14. Recta tangente y una secante "cercana".

Como todo el mundo sabe, la recta que pasa por los puntos $(a, f(a))$ y $(a+h, f(a+h))$ tiene por ecuación

$$y = \frac{f(a+h) - f(a)}{(a+h) - a}(x-a) + f(a). \tag{1}$$

(Insisto en que esto debería saberlo todo el mundo). La idea de que "exista la recta tangente" es que esa expresión tenga sentido "para valores infinitamente pequeños" de h. Como $f(a)$ es un número bien concreto, lo único que hace falta es que esta otra expresión tenga sentido:

$$\frac{f(a+h) - f(a)}{(a+h) - a} = \frac{f(a+h) - f(a)}{h}$$

Es decir, que *el cociente del incremento de f respecto del incremento de x* sea un número real cuando ambos incrementos son *infinitesimales*. Para esto es suficiente con que el incremento de f sea *del orden* del incremento de x, cuando este es muy pequeño, y que el cociente *tenga sentido*. En fin, de manera formal clásica:

DEFINICIÓN 18. Se dice que $f : X \to \mathbb{R}$ es derivable en un punto de la adherencia de X, $a \in \bar{X}$ si existe el siguiente límite

$$\lim_{x \in X, x \to a} \frac{f(x) - f(a)}{(x-a)}.$$

Cuando haya un entorno de a en X, se puede escribir

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Al valor de dicho límite se le llama *derivada de f en a* y se denota $f'(a)$ o bien $\frac{df}{dx}(a)$. Se dice que f es derivable en X si es derivable en todos los puntos de X.

Esta es la definición *newtoniana de derivabilidad.* Es mucho más fácil enunciarla como lo haría Leibniz: f es derivable si existe un infinitésimo d tal que

$$\frac{f(a+h)-f(a)}{h} = k + \mathrm{d}(h)$$

donde $k \in \mathbb{R}$ es independiente de d y dy es un infinitésimo cuando d lo es. Nótese que k no puede ser infinito, pues es un número: $k = f'(a)$.

Como se ve en ambas definiciones, la *inclinación* de la secante *infinitamente próxima* se mide escribiendo la ecuación de la secante como arriba, $y = kx + b$ y calculando la a (pues la b se deduce del punto $(a, f(a))$). Esto hace que, *si existe una única recta tangente*, entonces su pendiente a viene definida por el valor $k = f'(a)$. Sin embargo, como se ve en el ejemplo de $\sqrt{|x|}$, es posible que una función tenga una recta tangente en un punto *pero no tenga derivada* (esto ocurre cuando la tangente es *vertical*), porque la ecuación de una recta vertical es de la forma $x = c$, y nunca de la forma $y = kx + c$. Así pues:

Una función f es derivable en a
si su grafo tiene una sola tangente en $(a, f(a))$
y esta no es vertical.

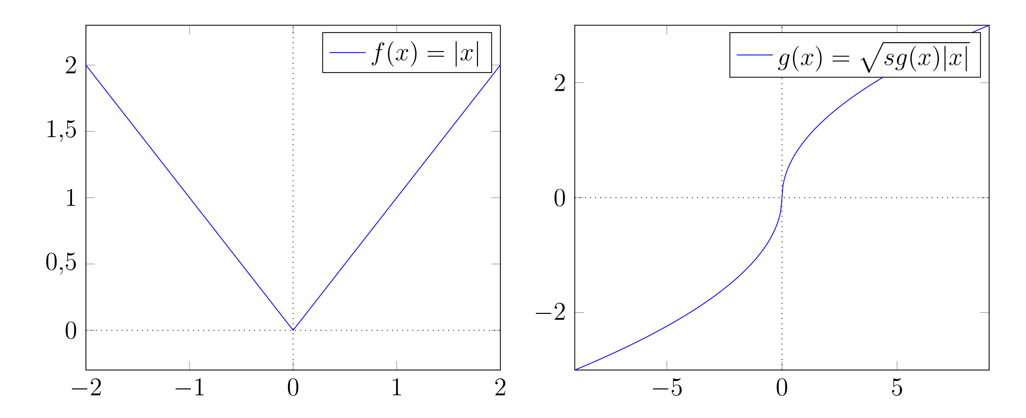


FIGURA 15. La función valor absoluto $|x|$ no es derivable porque “llega con dos tangentes” a $x = 0$. La función $sg(x)\sqrt{|x|}$ ($sg(x)$ es el signo de x) no lo es porque “es vertical” en $x = 0$.

Vamos a “demostrar” una propiedad esencial de las funciones derivables:

LEMA 4. *Si $f : X \to \mathbb{R}$ es derivable en a, entonces es continua en a.*

Demostración. Podríamos liarnos con ε y δ, pero es mucho más fácil usar infinitésimos. Que sea derivable en a quiere decir que existe $k \in \mathbb{R}$ tal que

$$\frac{f(a+h)-f(a)}{h} = k + \mathrm{d}(h)$$

donde d es un infinitésimo. Pero despejando, queda

$$f(a+h) = f(a) + (k+\mathrm{d}(h))h = f(a) + \mathrm{d}_1(h)$$

(donde $d_1(h)$ es un infinitésimo, obviamente), es decir, que $f(a+h)$ es $f(a)$ más un infinitésimo, que es justamente la noción de continuidad según la Definición 15. □

Esto significa que *ser derivable es más fuerte que ser continua*. La derivabilidad es lo que uno espera de los procesos físicos no traumáticos de magnitud no cuántica (los procesos cuánticos pueden ser discretos, no continuos, y requieren otro tipo de estudio). El interés de todo el resto de este tema (y de estudiar funciones varias veces derivables, etc.) reside precisamente en que *permite conocer con detalle los procesos físicos más habituales*. Existen procesos físicos que no se describen mediante funciones derivables, o incluso continuas, pero esos casos *singulares* no hacen que *el caso general deje de ser interesante*. De hecho, conocer el caso general es lo más útil (sirve más saber conducir por carretera que saber pilotar un Fórmula 1, aunque esto último sea mucho más excitante). Ha de entenderse todo lo que sigue como una serie de resultados que enriquecen el conocimiento de los *procesos normales*: las funciones varias veces derivables son las "normales", sus gráficas representan procesos "normales", y por eso conocer su *comportamiento* es útil.

Definición 19. Si una función $f : X \to \mathbb{R}$ es derivable en todos los puntos de su dominio, entonces se denomina *función derivada* de f, f' a la función que en cada punto x de X vale $f'(x)$.

6.1. Propiedades aritméticas de la derivada. Las propiedades básicas de la derivada se deducen muy fácilmente utilizando razonamientos con cantidades infinitamente pequeñas. Antes de nada:

Lema 5. *Si $f(x) = c$ en un intervalo (a, b), entonces $f'(x) = 0$. Si $f(x) = x$ en un intervalo (a, b), entonces $f'(x) = 1$. Es decir, $(c)' = 0$ y $(x)' = 1$.*

Demostración. La derivada es (si existe) el límite de:

$$\frac{f(x+h)-f(x)}{h}$$

en el primer caso, $f(x) = c$ siempre, así que queda

$$\frac{c-c}{h} = \frac{0}{h} = 0.$$

En el segundo caso, $f(x) = x$, así que

$$\frac{x+h-x}{h} = \frac{h}{h} = 1.$$

$\square$

De la misma manera se deducen las siguientes propiedades (suponiendo que f y g son funciones derivables en un punto x, etc…):

Suma: $(f+g)'(x) = f'(x) + g'(x)$.

Producto: $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.

Inversa: Si $f : (a,b) \to (c,d)$ es biyectiva y $f'(x) \neq 0$ y $f^{-1} : (c,d) \to (a,b)$ es su inversa, entonces $f^{-1}(f(x))' = \frac{1}{f'(x)}$.

División: Si $g(x) \neq 0$, entonces

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g(x)'}{g(x)^2}.$$

Potenciación: $(x^p)' = rx^{p-1}$, para $n \in \mathbb{Q}$ (fácil hacerlo para $p \in \mathbb{N}$, usar la función inversa para $1/p$ y usar otra vez la de $\mathbb{N}$ para p/q).

Composición: *o regla de la cadena*: $(f(g(x))' = g'(x)f'(g(x))$.

Vamos a demostrar solo unas de ellas.

Para la división, se tiene la siguiente cadena de igualdades

$$\frac{\frac{f}{g}(x+h) - \frac{f}{g}(x)}{h} = \frac{g(x)f(x+h) - g(x+h)f(x)}{g(x+h)g(x)h} =$$

$$\frac{g(x)(f(x) + f'(x)h + \mathrm{d}_1(h)h) - f(x)(g(x) + g'(x)h + \mathrm{d}_2(h)h)}{g(x)(g(x) + g'(x)h + \mathrm{d}_3(h))h} h =$$

$$\frac{f'(x)g(x)\mathrm{d} + g(x)h\mathrm{d}_1(h) - f(x)g'(x)h - f(x)h\mathrm{d}_2(h)}{(g(x)^2 + g'(x)h + \mathrm{d}_3(h))h} =$$

$$\frac{(g(x)f'(x) - f(x)g'(x) + \mathrm{d}_3(h))h}{(g(x)^2 + \mathrm{d}_4(h))h} = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2} + \mathrm{d}_5(h)$$

donde todos los d_i son infinitésimos. De aquí se deduce el resultado enunciado (hemos hecho uso de una igualdad $1/(a+\mathrm{d}(h)) = 1/a + \mathrm{d}_1(h)$, que es sencilla y se ha de conocer.

Para la *regla de la cadena*:

$$\frac{f(g(x+h)) - f(g(x))}{h} =$$

$$\frac{f(g(x) + g'(x)h + \mathrm{d}_1(h)h) - f(g(x))}{h} =$$

$$\frac{f(g(x)) + f'(g(x))g'(x)h + \mathrm{d}_1(h)\mathrm{d}_2(h)h - f(g(x))}{h} =$$

$$\frac{g'(x)f'(g(x))h + \mathrm{d}_2(h)h}{h} = g'(x)f'(g(x)) + \mathrm{d}_3(h)$$

donde todos los d_i son infinitésimos. De esta cadena de igualdades se deduce la regla de la cadena.

De aquí se deduce muy fácilmente el cálculo de la derivada de un polinomio, por ejemplo.

6.2. Derivadas de las funciones "típicas". ¿Cómo se calculan las derivadas del seno y del coseno? En realidad, esta pregunta es errónea: la función seno y la función coseno *se definen precisamente en función de sus derivadas.*

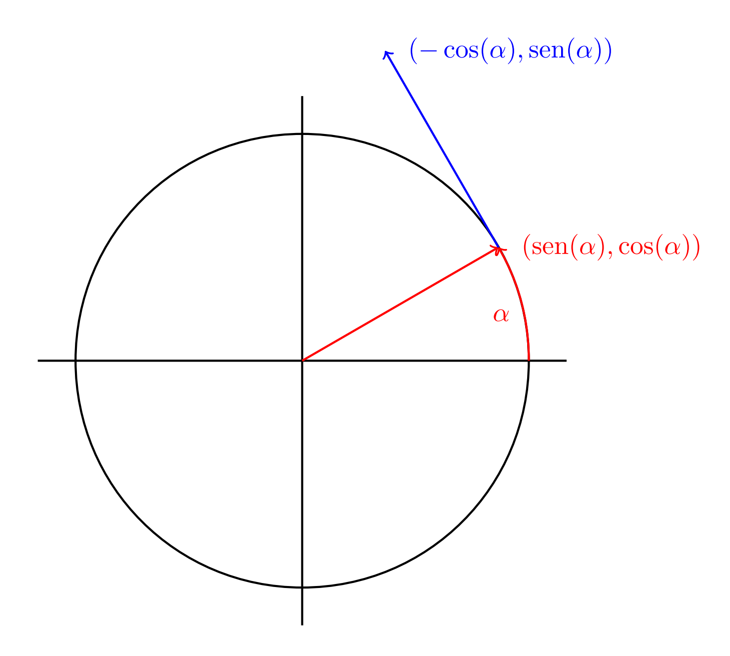


FIGURA 16. El seno y el coseno de α: la circunferencia unidad se recorre como $(\mathrm{sen}(\alpha), \cos(\alpha))$ a velocidad 1, así que el vector velocidad es $(-\cos(\alpha), \mathrm{sen}(\alpha))$, pues es *perpendicular* al radio.

La Figura 16 muestra gráficamente la definición de las funciones seno y coseno: son las que *parametrizan la circunferencia de radio* 1, de manera que a un ángulo α se le asigna $(\mathrm{sen}(\alpha), \cos(\alpha))$ de modo que el vector velocidad de dicha trayectoria tenga también módulo 1 y, al ser perpendicular al radio (esto es un principio físico), ha de ser $(-\cos(\alpha), \mathrm{sen}(\alpha))$. Pero el vector velocidad (y esto es otro principio físico) se calcula derivando coordenada a coordenada, así que al final nos queda que

$$\mathrm{sen}(x)' = \cos(x),\ \cos(x)' = -\,\mathrm{sen}(x),$$

que sabíais hace ya tiempo.

La derivada de la función exponencial es todavía más sencilla, pues la función exponencial se define como la única cuya derivada es ella misma:

$$(e^x)' = e^x,\ \text{ con } e^0 = 1.$$

De aquí sale la derivada del logaritmo. Como sabemos que $\log(e^x) = x$, tiene que ser

$$(e^x)'(\log(e^x))' = 1,$$

de donde, como la derivada de la exponencial es la exponencial,

$$(\log(e^x))' = \frac{1}{e^x},$$

es decir, que para todo número $x > 0$,

$$(\log(x))' = \frac{1}{x}.$$

6.3. La derivada logarítmica. No voy a insistir en esto, pero por la regla de la cadena

$$(\log(f(x)))' = \frac{f'(x)}{f(x)}$$

Así que si $f(x)$ es *difícil de derivar*, puede que sea más fácil derivar su logaritmo. Por ejemplo:

$$(x^x)' = ?$$

Tomando logaritmos, queda

$$(\log(x^x))' = (x\log(x))' = \log(x) + 1$$

y, por la regla de la cadena, es

$$(\log(x^x))' = \frac{(x^x)'}{x^x},$$

por lo tanto

$$(x^x)' = x^x(\log(x) + 1).$$

6.4. La recta tangente. El cálculo de los parámetros de la ecuación (1) (la de la "recta tangente" en $(a, f(a))$ a la gráfica de f) era uno de los objetivos a la hora de estudiar la derivada de una función de variable real. De todos los argumentos previos, se deduce que, si f es derivable en a, entonces la recta tangente a la gráfica de f en el punto $(a, f(a))$ es:

$$y = f'(a)(x - a) + f(a),$$

que es una expresión que *hay que saberse de memoria*. Bueno, realmente no hace falta saber nada de memoria: la recta tangente

- Tiene pendiente $f'(a)$ (esta es la definición de derivada).
- Pasa por el punto $(a, f(a))$ (pues es tangente a la gráfica).

Estas dos condiciones solo las cumple una recta... y esta hay que saber escribirla.

6.5. Órdenes de *magnitud* I. Regla de L'Hôpital. Los órdenes de magnitud de dos funciones se comparan dividiéndolas (no restándolas). Si $f(x)/g(x)$ es mayor que uno desde un momento en adelante (o cerca de un punto), sabemos que f tiene un orden de magnitud *al menos* como el de g. Si es cero, es que f es mucho más pequeña que g. A partir de ahora vamos a comparar funciones "que se acercan a cero" en un punto a, no necesariamente los infinitésimos del tipo $\mathrm{d}(h)$, que son los que hemos usado hasta ahora. La noción es esencialmente la misma.

DEFINICIÓN 20. Sea $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ un número real o infinito y sea f una funcion definida en un entorno de a. Se dice que f es un *infinitésimo* en a si $\lim_{x\to a} f(x) = 0$.

Es evidente que la suma, resta *y producto* de infinitésimos es un infinitésimo. También el producto de un infinitésimo por un número e incluso por una función *acotada* en un entorno de a.

Para comparar el orden de dos infinitésimos se utiliza el cociente:

DEFINICIÓN 21. Sean f y g dos infinitésimos en a y llámese l a $l = \lim_{x\to a} \frac{f(x)}{g(x)}$, que se supone que existe (o es $\pm\infty$). Se dice que:

- f y g son equivalentes si $l = 1$. Se escribe $f \equiv g$.
- f y g son del mismo orden si $l \in \mathbb{R}$ y $l \neq 0, 1$.
- f es de orden superior a g si $l = 0$. Se escribe $f = o(g)$.
- g es de orden superior a f si $l = \pm\infty$. Se escribe $g = o(f)$.

Con los infinitos se tiene una colección análoga de términos

DEFINICIÓN 22. Sea $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ un número real o $\pm\infty$ y sea f una función definida en un entorno de a. Se dice que f es un *infinito* en a si $\lim_{x\to a} f(x) = \pm\infty$.

A diferencia de los infinitésimos, con los infinitos solo se puede asegurar que el producto de un infinito por otro es un infinito y que el producto de un infinito por una función *cuya inversa está acotada cerca de a* es un infinito (es decir, el producto por una función $g(x)$ que tal que $|g(x)| > M$ para cierto M en un entorno de a).

Los infinitos también se comparan:

DEFINICIÓN 23. Sean f y g dos infinitos en a. Sea $l = \lim_{x\to a} \frac{f(x)}{g(x)}$, que se supone que existe (o es $\pm\infty$. Se dice que:

- f y g son equivalentes si $l = 1$. Se escribe $f \equiv g$.
- f y g son del mismo orden si $l \in \mathbb{R}$ y $l \neq 0, 1$.
- f es de orden superior a g si $l = \pm\infty$. Se escribe $f = O(g)$.
- g es de orden superior a f si $l = 0$. Se escribe $g = O(f)$.

NOTA 1. Esta notación, como la de la o, *no es universal*. Hemos utilizado, para conveniencia del alumno, la misma que en el libro de los profesores de la asignatura en el curso 2011/2012.

Está claro que habrá problemas a la hora de calcular estos límites, pues aparecerán siempre indeterminaciones del tipo $\frac{0}{0}$ o $\frac{\infty}{\infty}$. Se utilizará muchas veces la Regla de Lôpital:

TEOREMA 5 (Regla de Lôpital). *Sea $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ un número real o infinito. Sean $f(x)$ y $g(x)$ dos funciones definidas en un entorno de a y que son ambas infinitésimos o ambas infinitos en a. Supongamos que el siguiente límite existe:*

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = l \in \mathbb{R} \cup \{-\infty, +\infty\}.$$

Entonces el límite

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

también existe y vale l.

NOTA 2. Nótese —y esto es muy importante— la dirección de la implicación: si existe el límite del cociente de derivadas, existe el del cociente de funciones *y no necesariamente al revés.*

EJEMPLO 1. Sean $f(x) = x + \operatorname{sen}(x)$, $g(x) = x + \cos(x)$. Es relativamente sencillo probar que

$$\lim_{x\to \infty} \frac{f(x)}{g(x)} = 1.$$

Sin embargo, $f'(x) = 1+\cos(x)$, $g'(x) = 1-\operatorname{sen}(x)$. Estas dos funciones se hacen 0 en puntos distintos una infinidad de veces, así que el límite en infinito de $f'(x)/g'(x)$ no existe.

6.6. Los teoremas esenciales. Los problemas ingenieriles y físicos pueden ser, como ya dijimos, de varios tipos. Los dos básicos son:

- Resolver una ecuación (¿cuánto tiempo se tarda en esto?).
- *Optimizar* un proceso (¿cuál es la manera más rápida de hacer algo?).

El Teorema de Bolzano (Teorema 2) da una idea de que el primer tipo de problemas tiene solución si se conoce algo de información de la ecuación. De hecho, una manera burda pero funcional de resolver ecuaciones de manera aproximada es precisamente con la demostración clásica de dicho teorema por bipartición (esto lo estudiaréis en Métodos Numéricos).

Los problemas de optimización son por lo general muy complicados. Buscar máximos y mínimos puede ser una tarea difícil, tediosa y pesada. Hay múltiples algoritmos para localizar máximos y mínimos, pero todos ellos parten de un resultado gráfico básico: si una función $f(x)$ no tiene esquinas y a es, *localmente* un máximo o un mínimo, entonces la gráfica de f es *horizontal* en $(a, f(a))$. Las rectas horizontales tienen por ecuación $y = cte$.

DEFINICIÓN 24. Se dice que a es un *extremo local o relativo* de una función f si existe un intervalo $I = (a-\delta, a+\delta)$ en el que $f(a) \geq f(x)$ para todo x de I o bien $f(a) \leq f(x)$ para todo x de I. En el primer caso a es un *máximo local* y en el segundo un *mínimo local*. Si la desigualdad es $f(a) > f(x)$ para $x \neq a$, se habla de máximo local *estricto* (y lo análogo para el mínimo).

Nótese el adjetivo *local*: los extremos locales son extremos *cerca* de ellos, no necesariamente para toda la función:

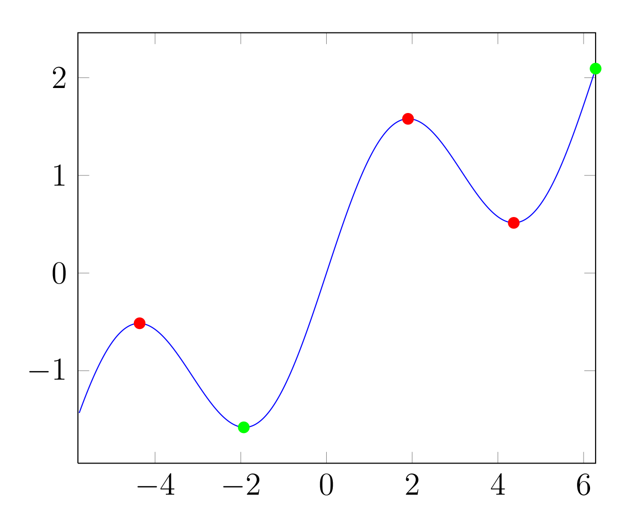


FIGURA 17. Extremos locales (rojo) y globales (verde) —los extremos globales también son locales, claro.

El resultado que acabamos de enunciar antes de la definición es el siguiente

LEMA 6. *Si $f : (a, b) \to \mathbb{R}$ es derivable en $c \in (a, b)$ y tiene un extremo local en c, entonces $f'(c) = 0$.*

DEMOSTRACIÓN. No hay más que ver la figura 17. Supongamos que f tiene un máximo en c. Entonces

$$f(c+h) = f(c) - \mathrm{d}(h)$$

donde $\mathrm{d}(h)$ es un infinitésimo *positivo*, independientemente de h. Como sabemos que es derivable, sabemos también que

$$\frac{f(c+h) - f(c)}{h} = f'(c) + \mathrm{d}_1(h)$$

Es decir,

$$f(c+h) - f(c) = (f'(c) + \mathrm{d}_1(h))h = f'(c)h + h\mathrm{d}_1(h).$$

Como $h\mathrm{d}_1(h)$ es *mucho más pequeño que* h, para que $f(c+h) - f(c)$ tenga signo negativo *independientemente de* h, solo puede ser $f'(c) = 0$ en el último miembro (si $f'(c) > 0$ entonces $f'(c)h + h\mathrm{d}_1(h)$ sería

positivo cerca de $h = 0$, y si $f'(c) < 0$ entonces sería negativo). Por tanto, tiene que ser

$$f'(c) = 0.$$

□

El razonamiento es *correcto*, aunque no estéis acostumbrados a pensar así. Insisto: así *piensan* los físicos, no con *épsilons y deltas* sino con "cantidades infinitamente pequeñas", y así se hizo el cálculo infinitesimal.

Nótese que el teorema enuncia una *condición necesaria*, pero no *suficiente*: hay puntos en los que la derivada puede ser 0 pero que no son extremos locales (ni globales).

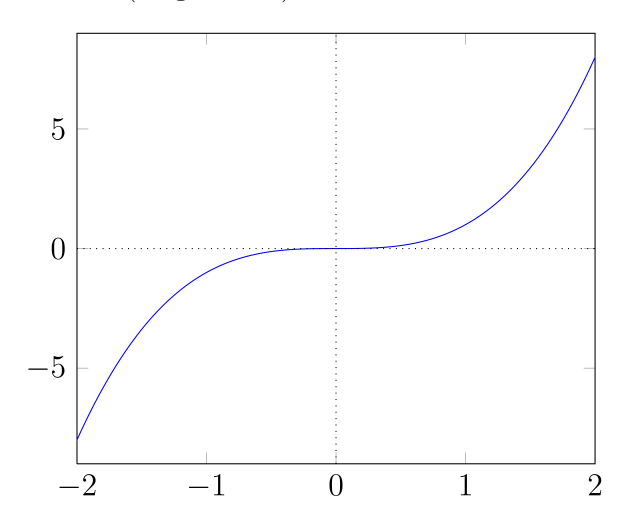


FIGURA 18. Un punto con derivada nula no extremo.

Del Teorema de Weierstrass (4) y de esta condición se deduce muy fácilmente el siguiente:

TEOREMA 6 (Teorema de Rolle). *Si $f : [a, b] \to \mathbb{R}$ es continua y es derivable en (a, b) y $f(a) = f(b)$ entonces hay un punto $c \in (a, b)$ tal que $f'(c) = 0$.*

Este resultado es la base del más importante de todo el Cálculo. Gráficamente quiere decir que si uno dibuja una función derivable (y por tanto continua) que vale lo mismo en los dos extremos de un intervalo, entonces la gráfica se hace horizontal en algún momento. Esto debería ser obvio. No por ello es *trivial*: este resultado es —llevado al extremo, que llamaremos el *Teorema del Valor Medio*— tan importante como el Polinomio de Taylor, que es lo más importante del Cálculo.

PRUEBA DEL TEOREMA DE ROLLE. La prueba es muy sencilla; hay dos posibilidades: o bien $f(x) = f(a) = f(b)$ en todos los puntos de $[a, b]$, o bien no es así. En el primer caso, la función es constante y su derivada es 0 en *todos* los puntos de (a, b). Fin. En el segundo caso,

por el Teorema de Weierstrass 4, la función tiene o bien un mínimo o bien un máximo *en el interior*: hay un punto $c \in (a, b)$ tal que $f(c)$ es un extremo de f (y por tanto, es un extremo *local*, por ser c del interior (a, b)). Acabamos de probar en el Lema 6 que $f'(c) = 0$. □

Téngase en cuenta que puede haber solo un extremo local, no tiene por que haber un máximo y un mínimo locales en (a, b) (piénsese en $f(x) = x^2$ definida en $[-1, 1]$, por ejemplo).

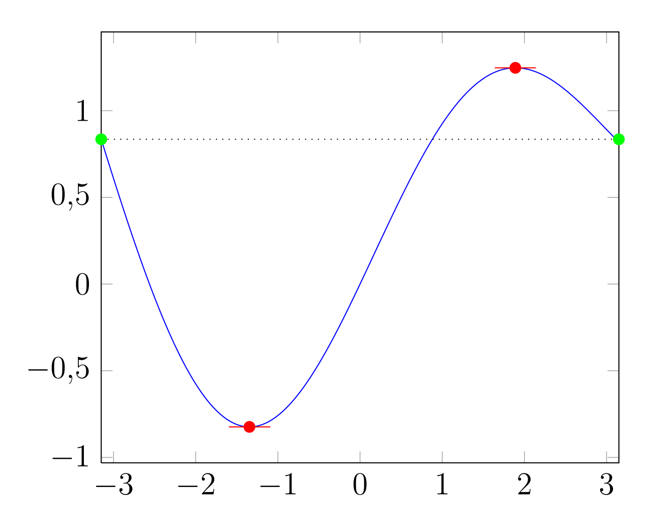


FIGURA 19. Puntos con derivada 0 (Teorema de Rolle)

6.7. El Teorema del Valor Medio. Del Teorema de Rolle se deduce, con un truco clásico (una trampa sintáctica, vamos a decir, porque hay que llamar a las cosas por su nombre), el Teorema del Valor Medio. Tiene varios enunciados. Primero damos el clásico, con la condición de igualdad. Es lo mismo que el Teorema de Rolle pero "inclinando la gráfica". Gráficamente: si una función es continua en $[a, b]$ y derivable en (a, b), entonces hay un punto $c \in (a, b)$ en que la derivada es paralela a la recta que pasa por los puntos $(a, f(a))$ y $(b, f(b))$.

TEOREMA 7. *Sea $f : [a, b] \to \mathbb{R}$ una función continua que es derivable en (a, b). Existe al menos un punto $c \in (a, b)$ en el que la derivada $f'(c)$ vale la pendiente del segmento que une $(a, f(a))$ y $(b, f(b))$:*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

DEMOSTRACIÓN. La prueba que sigue es un razonamiento geométrico clásico. Queremos convertir este enunciado en el del Teorema de Rolle (gráficamente es lo mismo salvo que en Rolle, la recta es "horizontal") y en este buscamos un punto en que la tangente sea paralela a la recta que une los extremos de la gráfica de f. Tendríamos que

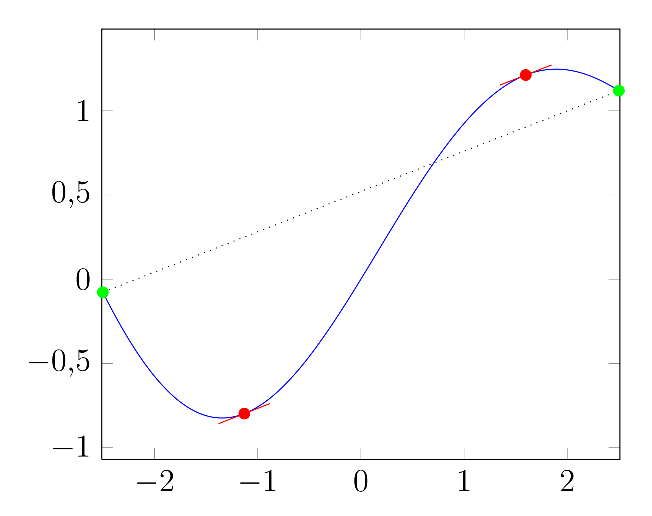


Figura 20. Teorema del Valor Medio. En este ejemplo hay *dos* puntos con tangente paralela a la recta que une los extremos. El teorema dice que *al menos* hay uno.

"llevar" $f(b)$ a la altura de $f(a)$, sin tocar $f(a)$. Nada más fácil. Vamos a definir una nueva función g que vale lo mismo que f en a y tiene una gráfica *similar* salvo que hacemos que $g(b)$ sea $f(a)$. Con un poco de esfuerzo…

$$g(x) = f(x) - (x-a)\frac{f(b)-f(a)}{b-a}$$

¿qué significa eso? Por un lado, si queremos que $g(x)$ *se parezca* a $f(x)$, empezamos haciendo que sea *igual*: $g(x) = f(x)$. Pero queremos deformarla *un poco* pero de tal manera que en a, se mantenga el valor: $g(a) = f(a)$. Esto es muy fácil, deformar significa sumar (o restar) algo. Pues restamos (se piensa más fácil restando):

$$g(x) = f(x) - \cdot$$

importante: queremos que lo que restamos *no toque el hecho de que* $g(a) = f(a)$, para ello nos aseguramos haciendo que ese resto valga 0 en a: pues multiplicamos por $(x-a)$:

$$g(x) = f(x) - (x-a)\cdots\cdot$$

Ahora ya se cumple, hagamos lo que hagamos al multiplicar, que $g(a) = f(a)$. ¿Podemos conseguir añadir un factor y que quede $g(b) = f(a)$ también? Aquí es donde la experiencia vale muchos puntos. ¿Cuánto vale $(x-a)$ en b? Pues $b-a$. Para empezar, queremos quitar este $(b-a)$ del problema: dividimos. Por tanto buscamos una expresión como

$$g(x) = f(x) - (x-a)\frac{\star}{b-a}$$

donde solo nos queda buscar el valor de $\star$. Tal como está la cosa, ¿cuánto vale ahora $g(b)$?:

$$g(b) = f(b) - (b-a)\frac{\star}{b-a} = f(b) - \star$$

y queremos que $g(b)$ sea $f(a)$. Es decir,

$$f(a) = f(b) - \star$$

Así que $\star = f(b) - f(a)$. De donde sale la expresión de $g(x)$. Está claro que $g(x)$ cumple las condiciones del Teorema de Rolle, así que hay un punto $c \in (a, b)$ en el que la derivada $g'(c)$ es 0. Pero…¿Cuál es la derivada de g?:

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b-a},$$

y como $g'(c) = 0$, queda

$$f'(c) = \frac{f(b) - f(a)}{b-a},$$

que es la pendiente de la recta que pasa por $(a, f(a))$ y $(b, f(b))$. □

El Teorema del Valor Medio *parece* que es un Teorema de igualdad (de hecho, tal como está enunciado, *es* una igualdad): existe un punto en que la tangente a la gráfica es paralela a la recta que une los extremos. Sin embargo, en *la vida real* se utiliza como un resultado de acotación, porque “se mira al revés”.

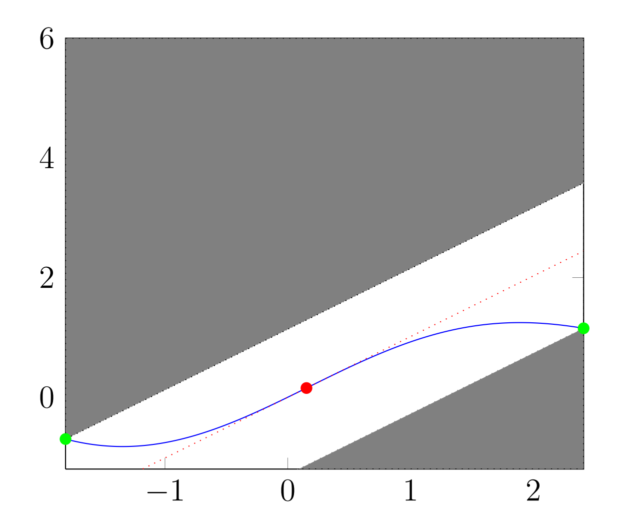


FIGURA 21. Teorema del Valor Medio como desigualdad. La pendiente más inclinada pone límites a los posibles valores de la función.

Observemos la Figura 21: la función (que es continua y derivable en el intervalo correspondiente) tiene un punto c (en rojo) en que la derivada es máxima (la curva tiene la mayor pendiente ahí). Por el Teorema del Valor Medio, si f tomara algún valor en la zona gris superior,

digamos $(x_1, f(x_1))$, habría un punto c_0 entre a y x_0 en que la derivada valdría

$$f'(c_0) = \frac{f(x_0) - f(a)}{x_0 - a},$$

pero eso quiere decir que la recta tangente a la gráfica de f en ese punto sería más pendiente que la recta roja punteada, y hemos dicho que esa es "la mayor pendiente" de f. Por un razonamiento análogo, la gráfica de f no puede pasar por la zona gris inferior. En resumen, el Teorema del Valor Medio nos dice que

Si podemos acotar la derivada de f en un intervalo, podemos acotar el incremento de f en ese intervalo.

Lo cual *no es nada nuevo*: si sabemos que Fernando Alonso ha ido como mucho a $300km/h$ durante un minuto, *es obvio* que podemos acotar el espacio que ha recorrido: como mucho $300km/h \cdot 1/60h = 5km$. Si sabemos que una barra de $1m$ de un material desconocido tiene al menos una densidad lineal de $2{,}75kg/m$, entonces la barra *al menos* pesa $1m \cdot 2{,}75kg/m = 2{,}75kg$.

El Teorema del Valor Medio como desigualdad no es más que una manera abstracta de enunciar esto.

Teorema 8 (del Valor Medio como desigualdad). *Supongamos que $f : [a, b] \to \mathbb{R}$ es continua en $[a, b]$ y es derivable en (a, b). Si existen números reales $m, M \in \mathbb{R}$ tales que $m \leq f'(x) \leq M$ en todo (a, b) entonces*

$$m(b - a) \leq f(b) - f(a) \leq M(b - a),$$

O, lo que es lo mismo:

$$m \leq \frac{f(b) - f(a)}{b - a} \leq M.$$

Es decir, acotar la derivada lleva a acotar el incremento de la función.

En castellano: si un motor rota entre $200rpm$ y $4500rpm$ y sabemos que está encendido durante $23s$, acótense las vueltas que puede haber dado. Fácil: como poco $200rpm \cdot 23/60m = 76{,}66vueltas$ y como mucho $1725vueltas$. Como todo el mundo sabe, las *revoluciones por minuto* son una derivada (en este caso, una derivada angular, pero derivada).

Este teorema nos servirá más adelante para... ¿cómo no? acotar integrales, claro.

6.8. Derivada de derivada etc... Si $f'(x)$ es la derivada de una función $f : \mathbb{R} \to \mathbb{R}$, se la puede considerar también como una función real de variable real, que puede a su vez ser derivable. La derivada da $f'(x)$ se llama *derivada segunda de f* y se denota $f''(x)$.

Y así sucesivamente, se definen la derivada tercera, la cuarta, la quinta, etc. La "general" se llama *derivada $n-$ésima.* A partir de $n = 3$, se suele indicar la derivada $n-$ésima con el número entre paréntesis (a

veces solo con el paréntesis cerrado), en lugar de una lista de comitas: $f^{(3)}(x), f^{(4)}(x), \dots, f^{(n)}(x)$.

6.8.1. Ejemplos. Calculemos algunas derivadas $n-$ésimas.

- La derivada $n-$ésima de la función exponencial e^x es, fácilmente, e^x, pues sabemos que $(e^x)' = e^x$, de manera que siempre sale lo mismo.
- La derivada $n-$ésima de la función x^m (donde m es, de momento, un número entero) se calcula poco a poco. La primera derivada es:
$$(x^m)' = mx^{m-1}.$$
"bajar el exponente y restarle uno". Si se deriva otra vez, el exponente ahora es uno menos y hay que restarle otro. Por tanto, si se deriva n veces, hay que multiplicar por $m(m-1)\dots(m-n+1)$ (cuidado con el $+1$), mientras que el exponente será $m-n$. Así que
$$(x^m)^{(n)} = m(m-1)\dots(m-n+1)x^{m-n}.$$
No hay que preocuparse de que $n > m$ pues en ese caso el coeficiente se hace cero y queda derivada $n-$ésima cero, como debe ser.
- La derivada $n-$ésima del logaritmo es un poquito más compleja. Se sabe que $(\log(x))' = \frac{1}{x} = x^{-1}$. Ahora se utiliza la derivada $n-1-$ésima que hemos calculado antes y ya está:
$$\left(x^{-1}\right)^{(n-1)} = (-1)(-2)\dots(-1-(n-1))x^{-1-(n-1)},$$
por tanto,
$$(\log(x))^{(n)} = (-1)(-2)\dots(-n)\frac{1}{x^n}.$$

Haremos más ejercicios, supongo.

6.9. Derivada de funciones implícitas. A veces una función viene definida de manera implícita, en lugar de "explícitamente". Un ejemplo bien conocido es la función que define la circunferencia de radio r centrada en $(0,0)$:
$$x^2 + y^2 = r^2.$$
La "gráfica" de esta función define, si se considera solo la parte "por encima del eje de las x", una función $y(x)$, de manera impícita. Pasando todos los términos al miembro de la izquierda, queda una igualdad a 0:
$$x^2 + y^2 - r^2 = 0.$$
Un problema que puede surgir es el de calcular la recta tangente a dicha curva implícita en un punto $(x, y(x))$ —por ejemplo, en $p = (r\frac{\sqrt{2}}{2}, r\frac{\sqrt{2}}{2})$. ¿Puede hacerse esto sin despejar? Sí, utilizando la regla de la cadena.

Como sabemos que y es una función de x (aunque no conozcamos su valor), podemos escribir

$$x^2 + y(x)^2 - r^2 = 0$$

y ahora, como la derivada de una constante es cero, si derivamos el miembro de la izquierda, debe dar cero. Usando la regla de la cadena:

$$2x + 2y(x)y'(x) = 0$$

De donde

$$y'(x) = -\frac{x}{y(x)}$$

y, en el punto p se tiene, por tanto

$$y'\left(\frac{r\sqrt{2}}{2}\right) = -1.$$

(lo cual es conocido: en ese punto la recta tangente a la circunferencia está inclinada 45 grados hacia abajo).

En general, siempre se calcula así. De hecho, puede también calcularse la derivada segunda, la tercera, etc... Siempre y cuando se cumplan unas condiciones mínimas que no vamos a explicar. A esto se le llama "derivación implícita". Cuando conozcamos el polinomio de Taylor, veremos que puede ser útil si se conoce una expresión implícita de una función, aunque no pueda despejarse. Haremos ejercicios de esto, pero no mucho más.

7. Taylor

La derivada, como se dice siempre, es la *pendiente de la recta tangente* a la gráfica de una curva. Dicho de otro modo, es la pendiente de la recta *que mejor aproxima* a una función en un punto. Es decir, si

$$y = k(x - a) + f(a)$$

es la ecuación de una recta que pasa por $(a, f(a))$, entonces esa recta es la tangente a la gráfica de f si y solo si $k = f'(a)$ (asumiendo que f es derivable en a, etc.).

Cabe preguntarse por las aproximaciones de segundo orden: ¿cuál es la curva de grado 2 que mejor aproxima a f cerca de a? Escribamos la ecuación de un polinomio de grado 2 que pase por $(a, f(a))$:

$$y = k_2(x - a)^2 + k_1(x - a) + f(a)$$

Queremos calcular k_1 y k_2 de manera que ese polinomio aproxime lo mejor posible a f cerca de a. Es decir, queremos (imitando el caso de la derivada) que

$$(2)\quad f(a + (x - a)) = f(a) + k_1(x - a) + k_2(x - a)^2 + (x - a)^2\mathrm{d}(x - a)$$

donde $\mathrm{d}(x-a)$ es un infinitésimo en a. Llamando $x-a=h$, se puede poner $(x-a)^2 \simeq h^2\mathrm{d}(h)$ con lo que queda

$$f(a+h) = k_1h + f(a) + h^2\mathrm{d}_1(h)$$

es decir, $k_1 = f'(a)$. Una vez calculado k_1, derivando la expresión (2), queda

$$f'(a+h) = 2k_2h + f'(a) + h\mathrm{d}_2(h)$$

pero si f' es derivable, esto no significa más que $k_2 = \frac{1}{2}f''(a)$.

Si en lugar de la aproximación de grado 2 usamos la de grado n, conseguimos un resultado análogo. Supongamos que $f(x)$ es derivable n veces y queremos aproximarla por un polinomio de grado n

$$y = a_n(x-a)^n + a_{n-1}(x-a)^{n-1} + \cdots + a_1(x-a) + a_0.$$

Llamando $(x-a)=h$, queda que

$$f(a+h) = a_0 + a_1h + h\mathrm{d}_1(h)$$

(para $\mathrm{d}_1(h)$ infinitesimal), de modo que $a_0 = f(a)$ y $a_1 = f'(a)$. Sabido esto, derivando sucesivamente se llega, tras n derivaciones, a que

$$f^{(n)}(a+h) = f^{(n-1)}(a) + (n)!a_nh + h\mathrm{d}_n(h)$$

para $\mathrm{d}_n(h)$ ifninitesimal, de donde se deduce que los coeficientes a_i son

$$a_0 = f(a), a_1 = f'(a), a_2 = \frac{1}{2}f''(a), \ldots, a_{n-1} = \frac{1}{(n-1)!}f^{(n-1)}(a),$$
$$a_n = \frac{1}{n!}f^{(n)}(a).$$

Así que, al final queda, poniendo $(x-a)$ en lugar de h

$$f(a+(x-a)) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots +$$
$$+ \cdots + \frac{1}{n!}f^{(n)}(a)(x-a)^n + \mathrm{d}_n(x-a)^n$$

con $dx_n(x-a)$ un infinitésimo.

De hecho, es posible acotar el error que se produce si se conoce algo sobre $f^{(n+1)}(a)$, como en el Teorema del Valor Medio. Pero vamos por partes.

Teorema 9 (Polinomio de Taylor). *Sea* $f : (a-\varepsilon, a+\varepsilon) :\longrightarrow \mathbb{R}$ *una función* n *veces derivable en* a. *Existe un polinomio* $P(x) = a_0 + a_1(x-a) + \cdots + a_n(x-a)^n$ *tal que*

$$\lim_{x\to a} \frac{f(x) - P(x)}{(x-a)^n} = 0.$$

De hecho, los coeficientes de $P(x)$ *son* $a_0 = f(a)$, $a_i = f^{(i)}(a)/i!$ *para* $i = 1, \ldots, n$.

A P_n se le llama *polinomio de Taylor de orden* n *de* f *en* a.

El siguiente resultado **es de examen**, no se puede decir más claro.

TEOREMA 10 (Primera cota del error). *Supongamos que f es derivable $n+1$ veces en $(a-\varepsilon, a+\varepsilon)$ y que P_n es su polinomio de Taylor de orden n en a. Si $|f^{(n+1)}(x)| < M$ en todos los puntos del intervalo $(a-\varepsilon, a+\varepsilon)$. Entonces*

$$|f(x) - P(x)| \leq \frac{1}{(n+1)!} M |x-a|^{n+1}$$

en dicho intervalo.

La prueba "formal newtoniana" de ambos resultados se hace por inducción (un método que para este caso es muy poco intuitivo). La prueba que hemos esbozado arriba es perfecta para la construcción del Polinomio de Taylor y, una vez construido, basta aplicar el Teorema del Valor Medio como cota para obtener el segundo.

7.1. Derivada implícita y Polinomio de Taylor. Quizás hagamos algún ejercicio sobre esto, pero depende del tiempo. Véase la sección 6.9 y repásense los detalles.

8. Propiedades dinámicas de las funciones

Hasta aquí hemos hecho estudios más bien teóricos de las funciones, aunque hemos también obtenido resultados muy interesantes Vamos ahora a estudiar las propiedades clásicas de las funciones, esencialmente todo lo relacionado con su *dinámica* (que es lo que *se ve* en la gráfica). En cierto modo podríamos titular esta sección *representación de funciones*, pero es mejor entender todo este trabajo como propiedades *dinámicas*.

8.1. Puntos de corte. Lo primero que interesa de una función f real de variable real son los puntos de corte con los ejes:

- Los puntos de corte con el eje OX son las raíces (soluciones) de la ecuación $f(x) = 0$.
- El (solo debería haber uno) punto de corte de f con el eje OY es el punto $(0, f(0))$.

Una *función* no puede tener más de un punto de corte con el eje OY (¿por qué?).

8.2. Extremos relativos. Recordemos la definición de *extremo relativo* (def. 24): supongamos que $f : [a, b] \to \mathbb{R}$ es continua y derivable en (a, b). Un *extremo relativo* de f en (a, b) (nótese que es en *el interior*) es un punto c en el que $f(c)$ es máximo localmente ó mínimo localmente (cerca de c). Recordemos que la derivada de f en un extremo relativo en (a, b) es 0 (insistimos en que debe ser en el interior).

Los valores extremales son interesantes. Si $f(x)$ es la energía de una partícula en la posición x, entonces un máximo es un punto de equilibrio *inestable* (¿por qué?) y un mínimo es un punto de equilibrio *estable* (¿por qué?). Como se ve, para calcular los valores extremales

se han de buscar primero los puntos de derivada 0, que se denominan *puntos críticos*. Una vez encontrados, hace falta discernir los máximos de los mínimos y de los que no son extremos.

LEMA 7. *Sea $c \in (a,b)$ un punto en que $f'(c) = 0$. Supongamos que las derivadas superiores son todas cero hasta la $k-1-$ésima. Entonces:*

- *Si k es impar, el punto* no es un extremo.
- *Si k es par, el punto es un máximo si $f^{(k)}(c) < 0$ y es un mínimo si $f^{(k)}(c) > 0$.*

DEMOSTRACIÓN. Esto es evidente. Sabemos, por el Teorema de Taylor, que

$$f(c+(x-c)) = f(c)+f'(c)(x-c)+\cdots+\frac{1}{k!}f^{(k)}(c)(x-c)^k+(x-c)^k\mathrm{d}(x-c)$$

donde $\mathrm{d}(x-c)$ es un infinitésimo. Pero hemos quedado en que todas las derivadas son 0 en c hasta la $k-1$. Por tanto, queda:

$$f(c+(x-c)) = f(c) + \frac{1}{k!}f^{(k)}(c)(x-c)^k + (x-c)^k\mathrm{d}(x-c).$$

Si k es impar, el factor $(x-c)^k$ cambia de signo al pasar por c y c no puede ser ni máximo ni mínimo. Si k es par, ese factor no cambia de signo y el término correspondiente *suma* si $f^{(k)}(c) > 0$ (así que c es un mínimo) y *resta* si es negativo (así que c es un máximo). □

Los puntos con derivada 0 que no son mínimos ni máximos son *puntos de inflexión*, que estudiaremos más adelante.

8.3. Concavidad y convexidad. La función más sencilla es una recta (las constantes son rectas). Lo normal es *comparar* cualquier gráfica con una recta. La derivada de una función sirve para esa comparación *de manera local*, cerca del punto en que se deriva. Una comparación global es más compleja pero puede llevarse a cabo con una buena definición.

DEFINICIÓN 25. Sea $f : [a,b] \to \mathbb{R}$ una función continua. Se dice que f es *convexa* si para cualesquiera $c < d \in [a,b]$ se tiene que *la gráfica de f está por debajo del segmento que une $(c, f(c))$ y $(d, f(d))$*. Dicho con propiedad:

$$f(c+\lambda(d-c)) \leq f(c) + \lambda(f(d) - f(c)).$$

Si la desigualdad es de signo contrario, se dice que f es *cóncava en* $[a,b]$.

Cuando una gráfica pasa de se cóncava a ser convexa, se nota cierta *torsión* en el punto correspondiente:

DEFINICIÓN 26. Un *punto de inflexión* de una función f es un punto c tal que f es convexa a un lado de c y cóncava al otro.

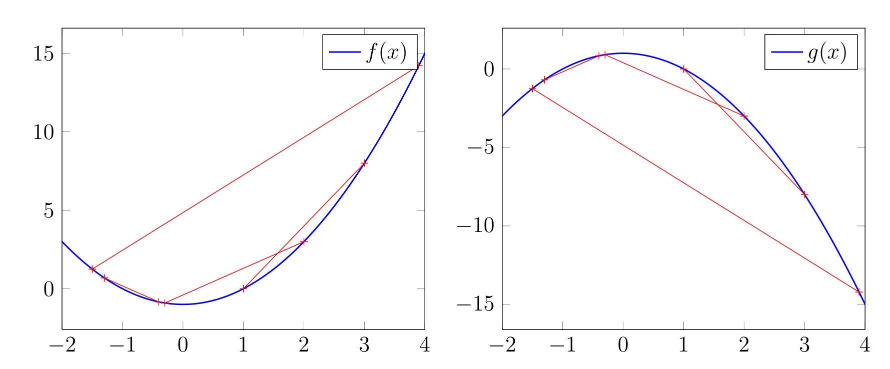


FIGURA 22. La función f es convexa: Todos los segmentos están "por encima" de la gráfica. La función g es cóncava.

La noción de concavidad y convexidad es muy relevante desde el punto de vista físico (y, en general, en cualquier contexto, un objeto convexo tiene unas propiedades físicas y geométricas especiales e interesantes). El caso es que *es posible detectar la convexidad mediante las derivadas de orden superior*. ¿Cómo se puede definir la convexidad de manera más local? Si uno le da suficientes vueltas, debería llegar a la conclusión de que:

Una función es convexa en $[a,b]$ si cada punto de la gráfica en el interior (a,b) es un mínimo local cuando se cambia el eje de las x por la recta que une dos puntos $(c, f(c))$ y $(d, f(d))$ con $c < d$.

Otra manera de verlo es:

LEMA 8. *Sea $f : [a,b] \to \mathbb{R}$. Si f es convexa en (a,b) y derivable dos veces en (a,b) entonces cualquier recta tangente a f en (a,b) queda* por debajo *de la gráfica de f. Si es cóncava en (a,b), entonces las rectas tangentes quedan por encima.*

Esta condición es muy sencilla de comprobar: Se está diciendo que

$$f(x_0 + (x - x_0)) = f(x) \geq f'(x_0)(x - x_0) + f(x_0)$$

cerca de cada punto x_0. Es decir:

$$f(x) - (f'(x_0)(x - x_0) + f(x_0)) \geq 0$$

pero está claro que el miembro de la izquierda vale 0 en x_0. Por tanto, se está exigiendo que el miembro de la izquierda *tenga un máximo* en x_0. Así pues, utilizando el Lema 7, se tiene que:

LEMA 9. *Sea $f : [a,b] \to \mathbb{R}$ una función derivable 2 veces y $c \in (a,b)$. Entonces:*

- *Si $f''(c) > 0$, entonces f es convexa en un entorno de c.*
- *Si $f''(c) < 0$, entonces f es cóncava en un entorno de c.*
- *Si $f''(c) = 0$, entonces no se sabe.*

Definición 27. Un *punto de inflexión* de $f : (a,b) \to \mathbb{R}$ es un punto $c \in (a,b)$ tal que f es cóncava a un lado de c y convexa al otro. Si f es derivable dos veces en c, entonces $f''(c) = 0$.

Pero no necesariamente ocurre que si la derivada segunda es 0, la función tiene ahí un punto de inflexión, hay que estudiar la concavidad y convexaidad a cada lado del punto.

8.4. Asíntotas. La gráfica de una función, una vez detectados los puntos de corte con los ejes, los extremos relativos y las zonas de concavidad y convexidad está prácticamente determinada, salvo el comportamiento *en el infinito*.

Por un lado está el infinito (positivo y negativo) del eje OX. Para estudiar cómo se comporta la función en esta dirección se calculará el límite cuando x tiende a $\pm\infty$ (es decir, dos límites distintos).

Para que la gráfica se vaya hacia infinito de modo paralelo al eje OY hace falta que haya un punto $c \in \mathbb{R}$ tal que o bien el límite por la derecha o bien el límite por la izquierda de $f(x)$ en c sea infinito.

De todos modos, el comportamiento en la dirección OX puede ser variado. Hay una caracterización que es bastante útil: las asíntotas.

Definición 28. Sea $f : \mathbb{R} \to \mathbb{R}$ una función y sea $r \equiv y = ax + b$ una recta. Se dice que r es una *asíntota* de f hacia $+\infty$ si

$$\lim_{x\to+\infty} f(x) - (ax - b) = 0.$$

(Se dice hacia $-\infty$ si ocurre lo mismo en $-\infty$). Si $a = 0$ entonces se habla de asíntota *horizontal*, si no, de asíntota *oblicua*.

Una *asíntota* es *una recta que se acerca mucho a la gráfica* (o más bien una recta a la que la gráfica se acerca mucho). ¿Cómo se sabe si una función tiene asíntotas en infinito?

Lema 10 (Cálculo de las asíntotas en infinito). *Sea $f : (A,\infty) \to \mathbb{R}$ una función. La recta $y = ax + b$ es una asíntota de f en $+\infty$ si y solo si se cumplen las dos condiciones siguientes:*

- $\lim_{x\to\infty} \frac{f(x)}{x} = a$.
- $\lim_{x\to\infty} f(x) - ax = b$.

(Lo mismo con $-\infty$ para las asintótas en $-\infty$).

La demostración de este lema es obvia, por la definición de asíntota.

Para calcular las asíntotas verticales hay que saber cuándo una función tiene límite infinito (p.ej. en los ceros del denominador es posible que ocurra...).

Además de acercarse, conviene saber si la función lo hace *por arriba* o *por debajo* de la asíntota. Para conocer esto es preciso conocer el signo de $f(x) - ax - b$: si es positivo desde un cierto momento, la asíntota queda por debajo de $f(x)$, si es negativo desde un cierto momento, la asíntota queda por encima. Podría cambiar de signo infinitas veces, en

cuyo caso la cortaría también un número infinito de veces. De todas maneras, es más sencillo lo siguiente:

LEMA 11. *Sea $f : \mathbb{R} \to \mathbb{R}$ una función. Supongamos que admite la asíntota $r \equiv y = ax + b$ en $\pm\infty$. Entonces:*

- *Si f es convexa a partir de un momento (bien por la izquierda, bien por la derecha), f queda por* encima *de r a partir de ese momento.*
- *Si f es cóncava a partir de un momento (bien por la izquierda, bien por la derecha), f queda por* debajo *de r a partir de ese momento.*

Como sabemos, para saber si f es convexa o cóncava, basta mirar el signo de la derivada segunda.

NOTA 3. Nótese que en una asíntota vertical, la función es convexa si va hacia $+\infty$ y cóncava si va hacia $-\infty$.

Esto es lo más que *se suele hacer*. Sin embargo, uno puede utilizar los *órdenes de infinitud* para, por ejemplo, dibujar varias gráficas en el mismo eje de coordenadas.

CAPÍTULO 3

Cálculo integral de una variable

Terminado el estudio *local* de las funciones reales de una variable real, especialmente el problema de la tangencia y la aproximación de funciones por polinomios (que son quizás los problemas esenciales del cálculo infinitesimal), se aborda clásicamente la teoría de integración, como la *operación dual* de la derivación.

En estos apuntes vamos a hacer una introducción más o menos rigurosa, sin preocuparnos demasiado por la formalidad, dando especial énfasis a la noción de integral como *masa* calculada a partir de una función de *densidad*. Pensamos que este (y no necesariamente el área bajo una curva) es el concepto clave con que los alumnos deben familiarizarse, pues es el concepto físico que van a encontrar. Con esta idea, utilizaremos la potencia de los teoremas del valor medio para acotar integrales (y por tanto acotar masas, espacios...), que pensamos que es lo más útil que se puede aprender, mucho más que las técnicas de integración, que hoy día son inútiles dada la ubicuidad de medios informáticos (cf. `wolframalpha.com`).

1. Integral de Riemann

Supongamos que se tiene una tabla como la 1, de velocidades instantáneas de un cohete en intervalos de $1s$, tal como las ha recogido el velocímetro. ¿Es posible hacerse una idea del espacio total recorrido por el cohete? (Asumiendo que la trayectoria es una línea recta perfecta, claro).

0	10
1	100
2	250
3	500
4	700
5	950
6	1100
7	1050
8	1090
9	1100

Cuadro 1. Velocidades de un cohete (en m/s).

Por supuesto, *algo* siempre se puede hacer, aunque sea inexacto. Lo más sencillo (que por supuesto es bastante impreciso, pero no deja de ser lo más sencillo) es suponer que la velocidad es constante en cada intervalo de segundo y simplemente calcular cuánto espacio habría recorrido a esa velocidad en ese tiempo y sumar todos los espacios:

$$10m/s \cdot 1s + 100m/s \cdot 1s + 250m/s \cdot 1s + \cdots + 1100m/s \cdot 1s = 6850m$$

casi siete kilómetros. Desde luego, esto es un cálculo bastante poco aproximado, pero algo tenemos.

Si el velocímetro tomara medidas con más frecuencia (cada centésima, por ejemplo, o cada milisegundo) tendríamos una tabla mucho más larga y *presumiblemente* el resultado final se acercaría más a la realidad.

Esa es la idea de Riemann, un poco más elaborada para que realmente tenga sentido la noción de que "uno se acerca" al valor real.

1.1. Particiones, suma inferior, suma superior. En Cálculo, más que buscar *resultados exactos* de manera directa, siempre se hacen las cosas *acotando*. La noción de límite clásica con ε y δ no es más que una manera técnica de afirmar que se pueden encontrar cotas cada vez más pequeñas del error de aproximación a un valor (el límite). Con la integral pasa algo similar.

Sea, de ahora en adelante, $[a, b]$ un intervalo compacto fijo (es decir, $a < b$ son dos números reales. Sea $f : [a, b] \to \mathbb{R}$ una función *acotada* (de momento solo ponemos esta condición, que relajaremos cuando hablemos de integrales impropias).

Definición 29. Una *partición* de $[a, b]$ es una familia finita de elementos de $[a, b]$ que empieza en a y termina en b:

$$a_1 = a < a_2 < a_3 < \ldots < a_n < a_{n+1} = b$$

Lo escribimos así para que el intervalo esté dividido en n subintervalos.

Como la función f es acotada en todo el intervalo $[a, b]$, en cada uno de los subintervalos $[a_i, a_{i+1}]$ y, por la completitud de $\mathbb{R}$, esto quiere decir que en cada subintervalo, f tiene un supremo y un ínfimo (no necesariamente un máximo y un mínimo, porque f no es continua necesariamente). Llamemos m_i al ínfimo en $[a_i, a_{i+1}]$ y M_i al supremo. Es decir, estamos seguros de que

$$m_i \leq f(x) \leq M_i \text{ en } [a_i, a_{i+1}]. \tag{3}$$

(así que la gráfica de f en ese intervalo está entre la banda horizontal de altura m_i y la de altura M_i).

Definición 30. Dada una partición P de $[a, b]$ en n subintervalos como arriba, se definen:

- La *suma superior* de f para P como

$$S_P(f) = \sum_{i=1}^{n} (a_{i+1} - a_i) M_i$$

- La *suma inferior* de f para P como

$$s_P(f) = \sum_{i=1}^{n} (a_{i+1} - a_i) m_i$$

que representan las *áreas* bajo las funciones escalonadas en esa partición.

Las *sumas superior e inferior* sirven como *cotas* del área entre el eje OX y la gráfica de f (poniendo en negativo lo que cae debajo del eje). Es decir, si se puede hablar de *área* de esa gráfica, entonces *tiene que estar entre esos dos valores.*

Una vez acotada la posible área, se procede como se hace siempre a *pasar al límite.* En este caso es algo más complicado, porque depende de la partición, etc. . . Vamos a hacerlo de manera directa y *a correr.*

Lema 12. *Las sumas superiores de f en $[a, b]$ están todas acotadas* inferiormente *y las sumas inferiores están todas acotadas superiormente.*

Demostración. Esto es evidente (o debería serlo). Es un buen ejercicio. □

Téngase en cuenta que —y esto es otro buen ejercicio— cuanto más *fina* es una partición (¿cómo se definiría esto?) la suma superior *decrece* y la suma inferior *crece.*

Con este resultado en la mano, sabemos que

- El conjunto de todas las sumas superiores de f en $[a, b]$ está acotado inferiormente.
- El conjunto de todas las sumas inferiores de f en $[a, b]$ esá acotad superiormente.

Por el axioma de los Intervalos Encajados (o por el *principio del supremo*) se deduce que

- El conjunto de las sumas superiores tiene un *ínfimo*. Llamémosle $S(f)$.
- El conjunto de las sumas inferiores tiene un máximo. Llamémosle $s(f)$.

La idea es decir que *una función tiene área si "el área por debajo es igual que el área por encima":*

Definición 31. Una función $f : [a, b] \to \mathbb{R}$ acotada se dice *integrable Riemann* si $S(f) = s(f)$. Si el ínfimo de las sumas superiores es igual al supremo de las sumas inferiores.

Existen funciones que *no son integrables Riemann* (la función característica de $\mathbb{Q}$, por ejemplo), pero la gran mayoría de las que uno se encuentra lo son:

LEMA 13. *Sea $f : [a,b] \to \mathbb{R}$ una función acotada. Si f es continua salvo en un número finito de puntos o si f es monótona* a trozos*, entonces f es integrable Riemann.*

NOTA 4. Una función $f : [a,b] \to \mathbb{R}$ es monótona a trozos si existe una partición de $[a,b]$, $a_1 = a, \dots, a_{n+1} = b$ tal que para todo i, f es monótona en (a_i, a_{i+1}).

DEFINICIÓN 32. Si $f : [a,b] \to \mathbb{R}$ es integrable Riemann, se define *la integral de f en $[a,b]$* como el supremo de las sumas inferiores o el ínfimo de las sumas superiores (que son iguales) y se denota

$$\int_a^b f(x)\, dx.$$

Donde x no es más que *el nombre de la variable* de integración. El factor dx es *fundamental*: indica que el valor $f(x)$ de la función de *densidad* se multiplica por la longitud del segmento "en el punto x" (una longitud infinitesimal) y queda $\frac{gr}{cm}\dot{c}m = gr$, masa.

La integral de una función es mucho más que el área entre la gráfica y el eje OX:

- Indica la *masa* del segmento $[a,b]$ si f denota una función de *densidad.*
- Indica la *carga total* del segmento $[a,b]$ si f denota una función de *densidad de carga.*
- Indica el *espacio recorrido* entre el instante a y el b si f es una función de *velocidad lineal* (el espacio recorrido en *línea recta*, claro).

Todos esos conceptos son los de integral y responden, claro a que *la función que se integra es la razón de lo que se calcula con respecto a la variable independiente*: la *densidad lineal* es la razón entre la masa y la longitud, la *densidad de carga lineal* es la razón entre la carga eléctrica y la longitud, la *velocidad* es la razón entre el espacio y el tiempo... Dicho de otro modo, la cantidad que se integra es, puntualmente, la *derivada* de la cantidad que se calcula.

2. Teoremas fundamentales

Antes de nada, debería ser muy sencillo comprobar (no lo vamos a hacer) que, si $f : [a,b] \to \mathbb{R}$ es una función acotada e integrable Riemann, entonces para cada $x \in [a,b]$, f es también integrable en $[a,x]$. Se puede por tanto definir la *función integral*:

DEFINICIÓN 33. Dada $f : [a, b] \to \mathbb{R}$ integrable Riemann, se define la función *integral de* f en $[a, b]$ como

$$F(x) = \int_a^x f(x)\, dx.$$

Por así decir, para cada x, la función $F(x)$ es la masa de la varilla de extremos a y x si la densidad lineal es $f(x)$. La relación entre estas dos funciones viene dada por el Teorema Fundamental del Cálculo Integral:

TEOREMA 11. *Sea* $f : [a, b] \to \mathbb{R}$ *una función continua (que, por tanto, es integrable Riemann) y sea* $F(x)$ *su integral* ($F(x) = \int_a^x f(t)\, dt$). *Se tiene que* $F(x)$ *es derivable en* (a, b) *y que*

$$\frac{dF}{dx}(c) = f(c)$$

la derivada de la función integral es la función que se integra.

Así pues, la idea *elemental* de que *la derivada de la masa es la densidad* es enunciable formalmente y es justamente el principio que acabamos de expresar. Más aun:

TEOREMA 12 (Regla de Barrow). *Supongamos que* $f : [a, b] \to \mathbb{R}$ *es integrable en* $[a, b]$ *y que* $F : [a, b] \to \mathbb{R}$ *es tal que* $F'(x) = f(x)$ *para cada* $x \in (a, b)$. *Entonces*

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

Esto es *más obvio de lo que parece*, aunque no sea obvio en absoluto. . .

Para diferenciar la *función integral* y la *integral*, se usa las siguiente nomenclatura:

Integral: Se denomina *integral definida* o *integral* de f (si $f : [a, b] \to \mathbb{R}$ es acotada e integrable en $[a, b]$) al *número*

$$\int_a^b f(x)\, dx$$

Primitiva: Se denomina *primitiva* de f (acotada e integrable en $[a, b]$) a *cualquier función* que cumpla $F'(x) = f(x)$ en cada punto de (a, b).

LEMA 14. *Supongamos que* $f : [a, b] \to \mathbb{X}$ *es acotada e integrable en* $[a, b]$ *y que* $F(x)$ *y* $G(x)$ *son dos primitivas de* f *en* $[a, b]$. *Entonces* $F(x) - G(x) = cte$.

Esto, junto con el Teorema Fundamental del Cálculo Integral quiere decir que si $F(x)$ es una primitiva de f en $[a, b]$ *donde* f *es acotada e integrable*, entonces existe una constante c tal que

$$F(x) = \int_a^x f(t)\, dt + c.$$

De ahí que se "diga" que *la primitiva de f es la integral más una constante arbitraria*. También se escribe:

$$\int f(x)\,dx$$

para denotar una primitiva de $f(x)$, sin especificar el intervalo de integración.

Otra igualdad que es *evidente* de todo lo dicho, es (en las condiciones adecuadas, etc):

$$\int f'(x)\,dx = f(x) + C,$$

que puede leerse *la integral es la operación inversa de la derivada.*

2.1. Propiedades de la integral *definida*. Fijamos para esta sección funciones $f, g : [a, b] \to \mathbb{R}$ *acotadas* e integrables. Entonces:

1. La integral es lineal: si $\lambda, \mu \in \mathbb{R}$:
$$\int_a^b \lambda f + \mu g\,dx = \lambda \int_a^b f(x)\,dx + \mu \int_a^b f(x)\,dx.$$
2. Se cumple la "desigualdad triangular" (aunque no tiene nada de triangular):
$$|\int_a^b f(x)\,dx| \leq \int_a^b |f(x)|\,dx$$
3. Si $f(x) \leq g(x)$ en $[a, b]$ y ambas son integrables, entonces
$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$
Este resultado es posiblemente el más útil en la práctica, pues permite, si se conoce alguna cota de la función, acotar la integral por arriba y por abajo.
4. Si $c \in [a, b]$ y f es acotada e integrable en $[a, b]$ entonces
$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$
(la integral es *aditiva* en el conjunto de integración).
5. Si $f : [a, b] \to \mathbb{R}$ es acotada e integrable y g es igual a f salvo en un número finito de puntos, entonces g también es integrable y
$$\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$$
6. La integral en un punto es cero:
$$\int_a^a f(x)\,dx = 0.$$

7. El siguiente resultado es consecuencia de los anteriores, pero es *importante*: si $f : [a,b] \to \mathbb{R}$ es convexa, acotada e integrable, entonces
$$\int_a^b f(x)\,dx \leq \frac{f(b)+f(a)}{2}(b-a),$$
pues el miembro de la derecha es el área que hay entre el eje OX y el segmento que une $(a, f(a))$ con $(b, f(b))$.⋆

Fixme: *dibujar el área aquí.*

Se podrían enumerar muchas más propiedades, pero las de arriba son quizás las más importantes y lo interesante es que, mediante ejercicios, el alumno se acostumbre a razonar *acotando* y estimando el valor de una integral (que es, entre otras cosas, lo que terminará haciendo en la asignatura de Métodos Numéricos).

Un último comentario. Se podría formalizar lo que sigue, pero lo vamos a enunciar como un mero axioma:

NOTA 5. Si el intervalo de integración está *de mayor a menor*, entonces el valor de la integral se cambia de signo:
$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$
Es importante hacer notar que esto se hace *exclusivamente* para simplificar los cambios de variables y *porque* $\mathbb{R}$ está orientado. En este sentido, podría formalizarse esta fórmula hablando de formas de volumen, etc... pero no vamos a liar al lector: tómelo como un axioma *que funciona.*

Para terminar, la siguiente notación es habitual:
$$H(x)\Big|_a^b = H(b) - H(a).$$

3. Teoremas de acotación

El resultado *dual* del Teorema del Valor Medio (Teoremas 8 y 7)

TEOREMA 13 (Teorema de la Media). *Sea $f : [a,b] \to \mathbb{R}$ una función continua (que, por tanto, es integrable Riemann). Existe $c \in (a,b)$ tal que*
$$(b-a)f(c) = \int_a^b f(x)\,dx.$$
De hecho, si m y M son el mínimo y el máximo de los valores de f en $[a,b]$, entonces se tienen las cotas:
$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$
Y cualquiera de las igualdades se da si y solo si f es constante.

De aquí se deduce que la integral de una función continua en $[a, b]$ y no negativa es no negativa (y, de hecho, es positiva si $f(c) > 0$ en algún punto de $[a, b]$).

4. Los dos "métodos" de integración

Uno de los problemas clásicos más *pesados* es calcular la primitiva de una función integrable. Para empezar, hay funciones cuyas primitivas se conocen por ser derivdas de funciones conocidas; estas se llaman *integrales inmediatas*. Unas cuantas se detallan en la Tabla 2. Nótese que en las integrales que especifican intervalos, *es esencial* que los intervalos estén especificados. Cualquier programa de integración calcula mucho más rápido que un humano, sin embargo.

$\int k\,dx = kx + C$	$\int x^\alpha\,dx = \frac{1}{\alpha+1}x^{\alpha+1} + C$
$\int \frac{1}{x}\,dx = \log\lvert x\rvert + C$	$\int c^x\,dx = \frac{1}{\log(c)}c^x + C\; c > 0$
$\int \mathrm{sen}(nx)\mathrm{d} = \frac{-1}{n}\cos(nx) + C$	$\int_{-1}^{x} \frac{1}{\sqrt{(1-x^2)}}\,dx = \mathrm{asen}(x) + C$
$\int \cos(nx)\mathrm{d} = \frac{1}{n}\,\mathrm{sen}(nx) + C$	$\int_{-1}^{x} \frac{-1}{\sqrt{1-x^2}}\,dx = \mathrm{acos}(x) + C$
$\int_{-\pi/2}^{x} \frac{1}{cos^2(x)} = \tan(x) + C$	$\int \frac{1}{x^2+1}\,dx = \mathrm{atan}(x) + C$

Cuadro 2. Algunas integrales inmediatas. Para las que dan el arcoseno y el arcocoseno, $x \in [-1, 1]$. Para la que da la tangente, $x \in [\pi/2, \pi/2]$.

Para calcular integrales y primitivas con `wolframalpha` no hay más que escribir

```
int(f(x),x,a,b)
```

que significa

$$\int_a^b f(x)\,dx$$

si se quiere calcular una primitiva, basta omitir los extremos del intervalo de integración.

Sin embargo, hay dos técnicas que se utilizan clásicamente para intentar reducir integrales a alguna integral *inmediata*: el *cambio de variable* y la *integración por partes*.

4.1. Cambio de variable. El cambio de variable en la integral es el resultado de *trasladar* la regla de la cadena (de la derivada) al cálculo integral. Recordamos que, si $g : [a, b] \to \mathbb{R}$ es derivable en (a, b)

y su imagen está contenida en $[c,d]$ $f:[c,d]\to\mathbb{R}$ es derivable en (c,d) entonces

$$f(g(x))' = g'(x)f'(g(x)).$$

Si en vez de derivar, se integra, como sabemos que *la integral es la operación inversa de la derivada*, queda algo *parecido*, aunque para enunciarlo con propiedad hay que tener más cuidado.

TEOREMA 14. *Sean $g:[a,b]\to\mathbb{R}$ y $f:[c,d]\to\mathbb{R}$ funciones acotadas e integrables y tales que $g([a,b])\subset[c,d]$. Supongamos además que g es derivable en (a,b) y sean $r=g(a), s=g(b)$. Entonces*

$$\int_{g(a)}^{g(b)} f(x)\,dx = \int_r^s f(x)\,dx = \int_a^b f(g(t))g'(t)\,dt$$

donde se sobreentiende (como se explicón en la Nota 5), que si $r>s$, se cambia el signo a la integral y se pone con los extremos bien ordenados.

La primera consecuencia obvia es que *las traslaciones de la variable* no afectan prácticamente a la integración: si $F(x)$ es la primitiva de f, entonces $F(x+c)$ es la primitiva de $f(x+c)$. Por ejemplo:

$$\int \operatorname{sen}(x+\omega)\,dx = -\cos(x+\omega)+C,$$

$$\int (x+a)^r\,dx = \frac{1}{r+1}(x+a)^{r+1} \text{ para } r\neq -1.$$

La siguiente es que las homotecias afectan a la integral *dividiendo*: si $F(x)$ es la primitiva de f, entonces, $\frac{1}{a}F(ax)$ es la primitiva de $f(ax)$. Por ejemplo:

$$\int \frac{1}{\sqrt{1-a^2x^2}}\,dx = \frac{1}{a}\operatorname{asen}(ax)+C, \quad \int \frac{1}{ax+b} = \frac{1}{a}\log|ax+b|$$

4.2. Integración por partes. Por otro lado, la fórmula de Leibniz para el producto de las derivadas:

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$$

permite simplificar algunas integrales, *escribiéndola así*:

$$f(x)g'(x) = (fg)'(x) - f'(x)g(x).$$

En concreto:

TEOREMA 15. *Sean f,g funciones acotadas integrables en $[a,b]$ y derivables en (a,b). Entonces*

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)\,dx.$$

Este resultado permite simplificar algunas integrales que contienen una parte *sencilla* de integrar y otra cuya derivada *simplifica* la expresión.

Ejemplos:

$$\int x \operatorname{sen}(x)\, dx.$$

no es inmediata, pero la función seno es sencilla de integrar y da una función también sencilla de integrar mientras que x, al derivarla queda 1. Así que (y aquí mostramos la manera *clásica* de escribir la integración por partes) se puede hacer:

$$\begin{array}{l|l} u = x & du = dx \\ dv = \operatorname{sen}(x)\, dx & v = -\cos(x) \end{array}$$

de donde

$$\int x \operatorname{sen}(x)\, dx = -x\cos(x) - \int \operatorname{sen}(x)\, dx,$$

la última integral es inmediata, así que

$$\int x \operatorname{sen}(x)\, dx = -x\cos(x) + \cos(x) = \cos(x)(1-x).$$

Si la integral fuera definida, habría que poner los extremos de integración por doquier *y tener mucho cuidado con los signos*.

4.2.1. El resto integral del desarrollo de Taylor. Recordemos que si $f : (a-\varepsilon, a+\varepsilon) \to \mathbb{R}$ es derivable $n+1$ veces en el intervalo, entonces, por la Fórmula de Taylor (Teorema 9), se puede escribir:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$

donde

$$|R_n(x)| \leq \frac{1}{n+1} M |x-a|^{n+1}.$$

El método de integración por partes permite, con un poco de ingenio, dar una fórumula *precisa* para el resto. En concreto,

$$R_n(x) = \int_a^x f^{(n+1)}(t) \frac{(x-t)^n}{n!}\, dt.$$

La prueba de esta fórmula es sencilla por inducción. Para el caso $n = 1$, se calcula

$$f(x) - f(a) = \int_a^x f'(t)\, dt$$

integrando por partes:

$$\begin{array}{l|l} f'(t) = u & f''(t)dt = du \\ dt = dv & t - x = v \end{array}$$

nótese que, por conveniencia, en lugar de tomar $\int dt = t$, tomamos $\int dt = t - x$, que también sirve. Pasando $f(a)$ al segundo miembro y utilizando la fórmula de integración por partes, queda

$$f(x) = f(a) + (t-x)f'(t)\Big|_a^x - \int_a^x (t-x)f''(t)\,dt$$

que, desarrollando, da

$$f(x) = f(x) + f'(a)(x-a) + \int_a^x (x-t)f''(t)\,dt$$

que es el resultado para $n = 1$. No hay más que *seguir utilizando la fórmula de integración por partes* para razonar por inducción.

Esta expresión del resto es *fundamental* porque permite concer la precisión de una fórmula. Por ejemplo: el desarrollo de Taylor de la exponencial es

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{1}{(n+1)!}\int_0^x (x-t)^n e^t\,dt$$

(puesto que la derivada $n-$ésima de e^x es e^x). Se sabe que $e^x < 3$ para $|x| \leq 1$, así que si $n = 10$, el error que se comete utilizando el polinomio de Taylor de grado 10 para calcular $e^1 = 1$, es menor o igual que:

$$r = \frac{1}{11!}\int_0^1 (1-t)^{10} e^t\,dt$$

pero $1 - t \leq 1$ si $t \in [0,1]$ y $e^t \leq 3$, por tanto:

$$|r| = r \leq \frac{3}{11!} = \frac{1}{13305600},$$

(aunque esta es una cota muy burda, pero da igual): es decir, la aproximación tiene al menos 6 dígitos decimales de precisión. Y es sencilla de calcular:

$$e \simeq 1 + \frac{1}{2} + \frac{1}{3!} + \cdots + \frac{1}{10!} \simeq 2{,}718281+$$

no está nada mal (de hecho hay otro dígito, un 8, que es preciso, pero no vamos a explicar por qué).

5. Integrales impropias

Hasta ahora hemos trabajado con funciones *acotadas* (la definición que hemos dado de integrabilidad solo se refiere a ellas) y con intervalos *acotados* (de la forma $[a,b]$ con $a, b \in \mathbb{R}$). Está claro que hay ocasiones en que o bien la función o bien el intervalo no está acotado pero tiene sentido preguntarse por la *integración.* Especialmente cuando, para simplificar, se modelan elementos de longitud o valor infinito. Por ejemplo: si una varilla infinita tiene densidad $e^{-1/x^2}\ g/cm$ en cada punto de coordenada x, ¿cuál es su masa? Si la velocidad de un proyectil se puede modelar como $\frac{1}{\sqrt{t}}$ para tiempo entre $t = 0s$ y $t = 1s$,

¿cuánto espacio recorre ese proyectil en ese tiempo? Ambos problemas pueden enunciarse, *de manera impropia* como dos integrales:

$$\int_{-\infty}^{\infty} e^{\frac{1}{x^2}}\,dx, \quad \int_0^1 \frac{1}{\sqrt{t}}\,dt$$

y se dice que *de manera impropia* porque en el primer caso el intervalo de integración no está acotado y en el segundo la función no está acotada. Podrían pasar las dos cosas a la vez, *ojo*.

Hay que tener en cuenta todos los casos posibles. Sean $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ dos números ó $\pm\infty$ y sea f una función $f : X \to \mathbb{R}$ donde X es (a, b) salvo un número *finito* de puntos $\{p_1, \dots, _n\}$ (ordenados de menor a mayor). Se dice que p_i es una *singularidad* de f si f no está acotada en ningún entorno $(p_i - \varepsilon, p_i + \varepsilon)$. La expresión

$$\int_a^b f(x)\,dx$$

se llama *integral impropia de primera especie de f en el intervalo* (a, b) si $a = -\infty$ o $b = +\infty$ (o ambos a la vez). Si ambos $a, b \in \mathbb{R}$ son números, entonces la expresión se llama *integral impropia de segunda especia de f en* $[a, b]$ si f tiene al menos una singularidad en $[a, b]$. Si no, se llama simplemente *integral de f* (aunque f no sea integrable, por abuso de lenguaje).

Definición 34. Sea $b \in \mathbb{R}$ un número y $f : (-\infty, b]$ una función *acotada*. Se dice que f admite integral impropia de primera especie si existe el límite

$$\lim_{K \to -\infty} \int_K^b f(x)\,dx$$

y, caso de que exista, se denota

$$\int_{-\infty}^b f(x)\,dx.$$

Si en lugar de $b \in \mathbb{R}$ se tiene $a \in \mathbb{R}$ y el intervalo es $[a, \infty)$ y f es *acotada en* $[a, \infty)$, se dice igual pero para este intervalo.

Cuando los dos extremos son infinito (uno positivo y otro negativo), *dividimos el intervalo en dos partes*:

Definición 35. Si $f : (-\infty, \infty) \to \mathbb{R}$ *acotada* admite integral impropia de primera especie

$$\int_{-\infty}^{\infty} f(x)\,dx$$

si para cualquier $c \in \mathbb{R}$ existen los límites

$$\lim_{K \to -\infty} \int_K^c f(x)\,dx, \quad \lim_{M \to \infty} \int_c^M f(x)\,dx$$

y la integral impropia es la suma de ambos.

Por otro lado, las integrales impropias de segunda especie se estudian *singularidad a singularidad*:

DEFINICIÓN 36. Sea $f : [a, b] \to \mathbb{R}$ una función tal que b es la única singularidad de f (es decir, f *no está acotada* en ningún entorno de b y esto solo pasa en b). Se dice que f admite integral impropia de segunda especie *por* b si existe el límite

$$\lim_{\varepsilon \to 0+} \int_a^{b-\varepsilon} f(x)\, dx.$$

Si la singularidad es a en vez de b, se procede de modo análogo. Si hay dos singularidades, se divide el intervalo en dos partes y se define como la suma de ambas integrales impropias.

Si $f : [a, b] \to \mathbb{R}$ tiene varias singularidades, se divide todo el intervalo $[a, b]$ en subintervalos $[a, p_2], \dots, [p_{n-1}, b]$ de manera que en cada intervalo haya singularidades en los extremos y se suman los valores de las integrales impropias según la definición 36.

Lo siguiente es tan sumamente importante que posiblemente caiga en todos los exámenes, por eso se recuadra:

De esta manera, es **absurdo** el siguiente cálculo:

$$\int_{-1}^{1} \frac{1}{x}\, dx = \log(|x|)\Big|_{-1}^{1} = 0 - 0 = 0.$$

porque el límite:

$$\lim_{\varepsilon \to 0} \int_{-1}^{\varepsilon} \frac{1}{x}\, dx = \lim_{\varepsilon \to 0} \log|x|\Big|_{-1}^{\varepsilon} = \lim_{\varepsilon \to 0} \log|\varepsilon| = -\infty$$

no existe como número real y por tanto la integral planteada *no está definida*, ni siquiera como integral impropia.

5.1. Convergencia de integrales impropias. Nos interesamos ahora por estudiar bajo que condiciones se puede afirmar que una integral impropia está o no definida. Estudiamos solamente (como es lógico) los casos en que hay una impropiedad (es decir, o bien un extremo es infinito o bien hay una singularidad). Comenzamos por las de primera especie y nos centramos en el límite $+\infty$.

TEOREMA 16. ⋆ *Sea $f : [a, \infty) \to \mathbb{R}$ una función real* positiva. *Supongamos que existe una constante M tal que* **Fixme:** *Comprobar*

$$f(x)x^{\alpha} \leq M.$$

Entonces: si $\alpha > 1$, la integral impropia

$$\int_a^{\infty} f(x)\, dx$$

existe (o converge*). Si $\alpha \leq 1$ entonces la integral impropia no existe (*no converge*).*

En concreto, si el límite

$$\lim_{x\to\infty} f(x)x^{\alpha}$$

existe y es finito para $\alpha > 1$*, la integral converge. Si ocurre para* $\alpha \leq 1$ y el límite no es 0*, la integral diverge.*

La condición de límite distinto de 0 en el segundo párrafo del teorema es esencial (límite 0 con $\alpha \geq 1$ no dice nada sobre la convergencia).

Para entender por qué el teorema es cierto, basta comprobar que la integral impropia

$$\int_1^{\infty} x^{-\alpha}\, dx$$

converge (existe) si $\alpha > 1$ y diverge (no existe) si $\alpha \leq 1$, lo cual se puede hacer calculando el límite aplicando la regla de Barrow a la primitiva.

Para las integrales impropias de segunda especie, pasa justo lo dual:

TEOREMA 17. *Sea* $f : (a, b] \to \mathbb{R}$ *una función real* positiva *con una singularidad en* a*. Supongamos que existe una constante* M *tal que*

$$f(x)(x-a)^{\alpha} \leq M.$$

Entonces: si $\alpha \geq 1$*, la integral impropia*

$$\int_a^b f(x)\, dx$$

no existe (o diverge*). Si* $\alpha < 1$ *entonces la integral impropia existe (*converge*).*

En concreto, si el límite

$$\lim_{x\to\infty} f(x)(x-a)^{\alpha}$$

existe y es finito para $\alpha < 1$*, la integral converge. Si ocurre para* $\alpha \geq 1$ y el límite no es 0*, la integral diverge.*

Para entender por qué el teorema es cierto, basta comprobar que la integral impropia

$$\int_0^{1} x^{-\alpha}\, dx$$

converge (existe) si $\alpha < 1$ y diverge (no existe) si $\alpha \geq 1$, lo cual se puede hacer calculando el límite aplicando la regla de Barrow a la primitiva.

6. Integrales paramétricas y derivadas

En distintos momentos de la vida (v.gr. cuando hay que utilizar la Transformada de Laplace o muchas otras transformadas) aparecen funciones que *se calculan como integrales*. La primitiva de una función es el ejemplo más sencillo:

$$F(x) = \int_a^x f(t)\, dt$$

viene definida como una integral diferente para cada valor de x. Hay otras expresiones, como

$$F(x) = \int_0^\infty e^{-t} t^{x-1}\, dt$$

en la que los extremos de integración no varían (de hecho, es una integral impropia) pero sí varía la función que se integra en cada x. En realidad esto es *más natural de lo que parece*, pero planteado así no es nada natural.

De todos modos, el hecho de que una función venga definida por una integral *hace posible estudiarla*: no es un fastidio, al contrario, de alguna manera permite acercarse a conocer las propiedades de dicha función. Y las propiedades, como hemos visto hasta ahora, se estudian, principalmente, *derivando*. ¿Cómo se derivan esas funciones? Con cuidado.

Comencemos con un ejemplo más o menos físico.

Ejemplo 2. Se tiene una plancha bidimensional de un material heterogéneo, que tiene forma rectangular salvo por el extremo superior, que se ha cortado siguiendo la función $f(x) = x^3$, donde x denota la disancia en horizontal al lado *izquierdo* de la plancha. Se sabe que la densidad lineal *en vertical* del material es proporcional a la distancia a la base, con constante de proporcionalidad k. Se van a fabricar varillas cortando en secciones verticales dicha planta. ¿Cómo variará la masa de las varillas?

El enunciado está preparado para escribir la masa de una varilla así:

$$m(t) = \int_0^{t^3} kx\, dx$$

El hecho de que el extremo superior de integración sea una función *no complica las cosas*. Llámese por ejemplo $g(t) = t^3$. Está claro que es

$$m(t) = \int_0^{g(t)} kx\, dx.$$

y no es más que

$$m(t) = F(g(t))$$

poniendo

$$F(u) = \int_0^u kx\, dx.$$

Si ahora queremos derivar $m(t)$, no hay más que aplicar la regla de la cadena:

$$m'(t) = g'(t)F'(g(t)) = 3t^2 F'(g(t))$$

pero $F(u)$ es una función definida como una integral, así que su derivada es el integrando: $F'(u) = ku$. Y como hay que evaluarla en $g(t) = t^3$,

queda
$$m'(t) = 3t^2F'(g(t)) = 3t^2k(t^3) = 3t^2kt^3 = 3kt^5.$$
A este resultado podíamos haber llegado sin tanto lío, calculando
$$m(t) = \int_0^{t^3} kx\,dx = \frac{k}{2}x^2\Big|_0^{t^3} = \frac{t^6}{2},$$
de donde
$$m'(t) = 3t^5.$$
Pero hay momentos en la vida en que *no se puede integrar tan fácilmente*... y sin embargo sí se puede derivar. En Wolframalpha es trivial:

```
D(int(k x,x,0,t^3),t)
```

Una función que *no se puede integrar* es e^{-x^2}, pero puede formar parte de integrales paramétricas:
$$F(t) = \int_{-\infty}^{\infty} t^2e^{-x^2}\,dx$$
(nótese cómo ahora lo que cambia no es el intervalo de integración, sino *la función que se integra*). ¿Se puede derivar $F(t)$? Resulta que sí (luego lo enunciaremos), pero no queda algo tan evidente como antes (y no necesariamente queda algo *calculable*). Si se deja x constante, la función integrando es
$$f(t) = t^2e^{-x^2}$$
cuya derivada es (la variable es t, la x es "constante"):
$$f'(t) = 2te^{-x^2}.$$
Pues bien, en este caso, se puede demostrar que
$$F'(t) = \int_{-\infty}^{\infty} 2te^{-x^2}\,dx.$$

Resumimos los resultados en los enunciados que siguen:

Sean $a(\lambda), b(\lambda)$ dos funciones reales de variable real definidas en un intervalo $X \subset \mathbb{R}$. Se considera una función $f(x,\lambda)$ que, para cada $\lambda \in X$ está definida y es acotada e integrable en $[a(\lambda), b(\lambda)]$. Llamemos
$$I(\lambda) = \int_{a(\lambda)}^{b(\lambda)} f(x,\lambda)\,dx$$
a la integral dependiente del parámetro λ. Queremos calcular su derivada.

Teorema 18. *En las condiciones anteriores, suponiendo que para cada x la función $f(x,\lambda)$ es derivable con respecto a λ y que la derivada es acotada e integrable en $[a(\lambda), b(\lambda)]$, se tiene que*
$$I'(\lambda) = b'(\lambda)f(b(\lambda),\lambda) - a'(\lambda)f(a(\lambda),\lambda) + \int_{a(\lambda)}^{b(\lambda)} \frac{\partial f(x,\lambda)}{\partial\lambda}\,dx$$

Nótese que si los extremos de integración son constantes, los dos primeros términos son nulos y desaparecen.

En general, si alguna de las integrales es impropia, no se puede hacer salvo que se cumplan fuertes condiciones de convergencia en las que no vamos a entrar. Sin embargo, si estas condiciones se dan, el enunciado sigue siendo cierto. Esto ese lo que ocurre en la transformada de Laplace de funciones *mucho más pequeñas que* e^x (subexponenciales).

CAPÍTULO 4

Sucesiones y series [... de funciones también]

Lo que nos interesa en este capítulo es acostumbrar al alumno a la notación de sucesiones y, quizás especialmente, conseguir que entienda las nociones de infinitos e infinitésimos equivalentes. Intentaremos que se haga algo de cargo de la utilidad del concepto de límite en la vida ¿real?.

El desarrollo de la sección sobre series de funciones se limita a las series de potencias, aunque procuraremos hablar de modo lo suficientemente general, al menos al introducir el concepto, que permita que entiendan en el futuro la noción de serie de Fourier, por ejemplo.

En realidad, casi las únicas aplicaciones de las sucesiones y todo esto es *verificar* que los cálculos aproximados que se realizan siempre (en la vida real) tienen una justificación e incluso su error puede acotarse con precisión (por eso, por ejemplo, se sabe que $e = 2{,}7182818+$ o que $\pi = 3{,}141592+$ con todas las cifras "exactas" (sin redondeos).

1. Sucesiones

La noción de sucesión refleja la idea de una *familia de números en evolución*. No es más que eso:

Definición 37. Una *sucesión* de elementos de un conjunto X es una aplicación $a_n : \mathbb{N} \to X$ del conjunto de números naturales en el conjunto X.

Esa definición significa literalmente "una familia de elementos de X ordenada como *el primero, el segundo, el tercero,...*". Cuando X es el conjunto de números reales, $X = \mathbb{R}$, se habla de sucesión de números reales o *sucesión numérica* (en estos apuntes, claro), o sin más, sucesión. Cuando se habla de los elementos de una sucesión en general, se habla del *término general*.

Ejemplo 3. Varios ejemplos:

- La sucesión *constante*: $c, c, c, c, \dots$, se escribe $a_n = c$. *No es un número*, es una familia de números iguales.
- La sucesión nula: $0, 0, 0, \dots$, $a_n = 0$.
- La sucesión de los números impares: $1, 3, 5, 7 \dots$, su término general puede escribirse $a_n = 2n - 1$ (si los naturales empiezan en 1, cosa "discutible").

- La sucesión de los pares: $0, 2, 4, 6, \dots$, puede escribirse $a_n = 2(n-1)$, o, si los naturales empiezan en 0, $a_n = 2n$. Pero hay que especificarlo antes de decir nada.
- La sucesión de los primos: $2, 3, 5, 7, 11, 13, 17, \dots$. No sé conoce una expresión simplificada para ella. Este hecho es uno de los pilares de la criptografía moderna.
- La sucesión $e^{\pi n}$ comienza con $e^0 = 1$, y luego ya los demás son todos raros: e^{π}, $e^{2\pi}$,etc…
- La sucesión de término general $1/n$ es muy importante en muchos casos.
- La sueción $1, 1, 2, 3, 5, 8, 13, \dots$ de números de Fibonacci no tiene una expresión sencilla, salvo que $a_n = a_{n-1} + a_{n-2}$ para $n > 2$ y $a_1 = 1, a_2 = 1$. Se dice que está definida *por recurrencia* (se *recurre (vuelve)* a los términos anteriores para calcular el presente).
- En la vida real, uno se encuentra con la sucesión de pagos de una hipoteca, que puede ser constante o puede ser variable. O con la sucesión de beneficios de unas acciones. O con la sucesión del capital acumulado por una inversión (digamos en acciones) a lo largo del tiempo (cada mes puede actualizarse, o cada día).

El ejemplo clave: el número e.

Ejemplo 4. Una empresa ofrece préstamos a un año a interés compuesto calculable en fracciones de distinto periodo. Por ejemplo, puede pedirse un préstamo a un interés fijo de un año, o a un interés fijo mensual, o a un interés fijo diario, o incluso con la amortización que se quiera. Si el interés anual es r, compruébese que el capital al cabo de un año en un préstamo que se divide en n unidades de $1/n$ años es

$$K = C\left(1 + \frac{r}{n}\right)^n$$

donde C es el capital inicial.

¿Qué ocurre cuando se divide el año en muchísimas partes? Resulta que uno tiene que preocuparse por la sucesión de término general

$$a_n = \left(1 + \frac{r}{n}\right)^n$$

y estudiar qué pasa si n es “gigantesco”…

Lo que ocurre es que esa sucesión se aproxima mucho (cada vez más y todo lo que uno quiera) a un número (que depende de r, claro): e^r. Si $r = 1$, queda e.

Por eso en los mercados de valores (especialmente en productos de muy alta frecuencia de inversión) en lugar de calcularse el interés por medio del polinomio de grado n, se asume que *el tiempo es continuo* y se calculan los intereses como

$$K = Ce^r,$$

que es el *límite* de la expresión de arriba cuando n es muy grande (imagínese que uno actuliza el capital cada segundo, por ejemplo, o cada milisegundo).

Las sucesiones $1/n$, $2 + \frac{(-1)^n}{n}$, $\left(1 + \frac{1}{n}\right)^n$ tienen algo en común que es difícil de expresar de otra manera que con la defininción de límite:

DEFINICIÓN 38. Una sucesión a_n se dice que tiene límite $l \in \mathbb{R}$ si dada cualquier distancia $\varepsilon > 0$, existe un índice n_0 tal que $|a_n - l| < \varepsilon$ para cualquier $n \geq n_0$. Se indica

$$\lim_{n\to\infty} a_n = l \text{ o bien } a_n \to l.$$

Es decir, una sucesión tiene límite l si cualquiera que sea la distancia elegida (por pequeña que sea), *todos* los términos a partir de un momento están más cerca de l que esa distancia.

Así pues, la sucesión $1, 0, 1, 0, 1, 0, \dots$ *no tiene límite*. La sucesión $1, 2, 3, 4 \dots$ tampoco, la $0, 0.1, 0.01, 0.001, \dots$ *sí*: ¿cuál? ¿por qué?

De todos modos, hay una diferencia entre $0, 1, 0, 1, \dots$ y $2, 4, 6, 8, \dots$: la primera *no va a ninguna parte* mientras que la segunda *se hace cada vez más grande*. Esto se diferencia con la noción de *límite infinito*:

DEFINICIÓN 39. Se dice que la sucesión a_n tiene por límite $+\infty$ si dado cualquier $M > 0$, a partir de un término n_0, todos los a_n son mayores que M. Para $-\infty$ se toma $M < 0$ y $a_n < M$.

Eso "normaliza" el *tan grande como se quiera* a partir de un momento. Aun así, la sucesión $1, 0, 2, 0, 3, 0, \dots$ no tiene límite ∞, obviamente, porque aunque muchos términos se hacen "muy grandes", no lo hacen *todos*.

1.1. Complejidad - importante. Algoritmos de ordenación o de búsqueda.

EJEMPLO 5. El problema de la ordenación. Se tiene una colección de n elementos (por ejemplo, n transacciones financieras, esto en un mercado de valores puede ser un número gigantesco) y hay que ordenarlas (p.ej. por fecha). Se pueden describir dos algoritmos de ordenación:

- **Burbuja:** Se compara el primero con el segundo y si están desordenados, se intercambian. Una vez hecho esto, se compara el segundo con el tercero... Así hasta llegar al último. En esta primera pasada, el último es el mayor. Se apunta como el mayor y se repite el proceso con todos menos el mayor...
- **Mezcla:** Si la lista tiene 0 ó 1 elementos, está ordenada. Si no, se divide la lista en dos partes más o menos iguales. Se ordena cada una de estas partes del mismo modo. Al final, se mezclan ordenadamente las dos listas. (¿alguien lo ha entendido?).

¿Cuál es el mejor de los dos algoritmos? La burbuja necesita unas n^2 comparaciones, la mezcla unas $n\log(n)$: ¿qué es mejor? Pues si sabemos que

$$\lim_{n\to\infty}\frac{n\log(n)}{n^2}=0,$$

esto quiere decir que el número de operaciones que requiere la mezcla es mucho menor que el que requiere la burbuja. Así que, *en principio*, es mejor la mezcla. **Discutir esto sería importante: ¿mejor por qué? ¿y si hay limitaciones de memoria, por ejemplo?**

1.2. Progresión aritmética y geométrica.

Definición 40. Una sucesión a_n se llama *progresión aritmética* si la diferencia de dos elementos sucesivos es una constante: $a_{n+1}-a_n=c$ para todo n.

Definición 41. Una sucesión a_n se llama *progresión geométrica* si el cociente entre dos elementos sucesivos es una constante: $a_{n+1}/a_n=c$ para todo n (o bien todos los elementos son 0).

Se puede calcular con una sencilla fórmula la suma de los k primeros términos de una progresión aritmética y el producto y la suma de los k primeros términos de una progresión geométrica. Son ejercicios interesantes, hasta cierto punto.

1.3. Álgebra de límites.

Teorema 19. *Si $a_n\to a$ y $f:\mathbb{R}\to\mathbb{R}$ es continua en a entonces*

$$\lim_{n\to\infty}f(a_n)=f(a).$$

Demostración. Obvio: esto es *exactamente* otra manera de definir la noción de *función continua*. □

De este resultado se deduce prácticamente todo lo que sigue.

Sean a_n y b_n dos sucesiones y supongamos que ambas tienen límite finito, $a_n\to a$ y $b_n\to b$.

$$(4)\quad \begin{gathered}(a_n+b_n)\to a+b,\ a_nb_n\to ab,\ \frac{a_n}{b_n}\to\frac{a}{b}\text{ si } b_n,b\neq 0,\\ a_n^{b_n}\to a^b \text{ si } a_n,a\neq 0, b_n\geq 0,\ \log(a_n)\to\log(a)\text{ si } a_n,a>0.\end{gathered}$$

1.4. Órdenes de magnitud. Muchas veces es más interesante la noción de orden de magnitud que la de límite, al igual que muchas veces es más interesante acotar que calcular con precisión un valor. Podríamos hacer el estudio todo junto (con sucesiones y límites de cocientes, etc. . .) pero para unificar la exposición, hablaremos de órdenes de infinitos y órdenes de infinitésimos.

1.4.1. Infinitésimos. Se dice que una sucesión a_n es un infinitésimo si $a_n \to 0$. Fijemos para este apartado dos infinitésimos a_n y b_n. Del mismo modo que para infinitésimos de funciones (Definición 20), decimos que a_n y b_n son

Equivalentes: si $\frac{a_n}{b_n} \to 1$.
Del mismo orden: si $\frac{a_n}{b_n} \to r \in \mathbb{R}$ y $r \neq 0$ (esto *incluye* que sean equivalentes).

Por otro lado, si $\frac{a_n}{b_n} \to 0$ se dice que a_n es de orden mayor que b_n; si $\frac{a_n}{b_n} \to \infty$, [obviamente] b_n es de orden mayor que a_n. Si no pasa ninguna de todas estas condiciones, son *incomparables*.

El *orden* de un infinitésimo indica *con que velocidad* se acerca a cero (relativamente a otro infinitésimo, claro). Si a_n es de mayor orden que b_n, es porque a_n es *muchísimo más pequeño* que b_n (siempre para $n >> 0$). Son equivalentes si, salvo pequeñas variaciones, van a la misma velocidad. Y son del mismo orden si las velocidades son comparables (el cociente "es un número real").

Por lo general, se escribe $a_n \sim o(b_n)$ si a_n es de orden mayor que b_n como infinitésimo.

1.4.2. Infinitos. Para infinitos se dan las mismas definiciones, que no son más que la traslación a este contexto de las anteriores. Supongamos que a_n y b_n tienen *ambas* límite ∞. Se dice que son

Equivalentes: si $\frac{a_n}{b_n} \to 1$.
Del mismo orden: Si $\frac{a_n}{b_n} \to r \in \mathbb{R}$ y $r \neq 0$ (esto *incluye* que sean equivalentes).

Por otro lado, si $\frac{a_n}{b_n} \to 0$ se dice que b_n es de orden mayor que a_n.

Si $\frac{a_n}{b_n} \to \infty$; [obviamente] a_n es de orden mayor que b_n. Si no pasa ninguna de todas estas cosas, son incomparables.

Téngase en cuenta que, como se están mirando las cosas "en la dirección contraria", la noción de *mayor orden* corresponde al que va *más rápido hacia* $+\infty$.

La notación clásica es $a_n \sim O(b_n)$ cuando a_n y b_n son del mismo orden. Aunque esto difiere. Desde luego, si son del mismo orden, esta notación es correcta.

Por lo general, interesa acotar el orden de magnitud de los infinitos, más que nada para hacerse una idea. El problema del tiempo polinomial para llevar a cabo un algoritmo.

1.5. Cálculo de límites: "trucos". Espero que quede claro.

Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador

o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites,

en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un

ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet. Los límites, en general, se calculan con un ordenador o usando Internet.
El siguiente comando en `wolframalpha.com`:

```
lim(sin(n)*log(n)/sqrt(n),n,inf)
```

calcula el límite de la sucesión $\frac{\operatorname{sen}(n)\log(n)}{\sqrt{n}}$, que es 0. Lo podéis probar con vuestro teléfono móvil.

Solo los límites de sucesiones que no se pueden escribir como funciones conocidas puede que requieran cálculos a mano previos a introducirlos en un ordenador.

Antes de seguir, generalizamos la noción de equivalencia a cualquier par de sucesiones:

DEFINICIÓN 42. Dos sucesiones a_n y b_n son *equivalentes* si

- O son ambas infinitos o infinitésimos equivalentes, o bien
- Existe el límite del cociente y es 1: $\lim \frac{a_n}{b_n} = 1$[1].

Como se ha visto antes (Teorema 19), las funciones continuas permiten sustituir el límite de una sucesión en el punto para calcular el límite de una *función de una sucesión*. De este modo, es obvio que:

1. Si $P(n)$ y $Q(n)$ son dos polinomios y sus términos de mayor grado son $p_m n^m$ y $q_r n^r$, entonces:

$$\lim \frac{P(n)}{Q(n)} = \lim \frac{p_m n^m}{q_r n^r} = l$$

   donde l es:
   - 0 si $m < r$,
   - p_m/q_r si $m = r$,
   - ∞ si $m > r$.

[1]Esto requiere que $b_n \neq 0$ para $n >> 0$, lo cual siempre se supondrá cuando se use esta noción

2. **Criterio del cociente**: Si existe $\lim \frac{a_{n+1}}{a_n}$ y a_n es de términos positivos, entonces
$$\lim \frac{a_{n+1}}{a_n} = \lim \sqrt[n]{a_n}$$
3. Si a_n y b_n son dos sucesiones *equivalentes* con límite l y $f(x)$ es una función continua en l, entonces
$$\lim f(a_n) = \lim f(b_n).$$
4. Si a_n y b_n son dos sucesiones equivalentes y c_n es otra sucesión, entonces: si uno de los miembros de cada igualdad existe, entonces existe el otro y se da la igualdad
$$\lim a_n c_n = \lim b_n c_n, \;\; \lim \frac{c_n}{a_n} = \lim \frac{c_n}{b_n}$$ [2]

1.5.1. Casos "indeterminados". Todos estos casos *deberían ser bien conocidos* y vamos a pasar por ellos rapidísimamente.

El caso $\frac{0}{0}$: se da, obviamente, cuando se divide un infinitésimo entre otro. Por lo general, todos estos se resuelven utilizando desarrollos de Taylor (u otros desarrollos) de las funciones que definen las sucesiones cerca de 0. Es decir, hace falta conocer el desarrollo de Taylor de las funciones más importantes.

Bueno, en realidad no hace falta conocerlos: lo que hace falta es saber cómo se calcula: derivando...

Y, por otro lado, tener clara la siguiente tabla de órdenes infinitesimales (donde $c > 1$), de izquierda a derecha van de menor a mayor orden (es decir, la división de uno de la izquierda por uno de la derecha tiene por límite 0):

(5)

$\frac{1}{\log(n)}$	$\frac{1}{n^c}$	$\frac{1}{c^n}$	$\frac{1}{n!}$	$\frac{1}{n^n}$

El Polinomio de Taylor es crucial para este caso:

Lema 15. *Si $f(x)$ continua y nula en 0 tiene desarrollo de MacLaurin $P(x) = p_k x^k + \dots$, con $p_k \neq 0$ (es decir, si su desarrollo no es nulo), entonces, para cualquier infinitésimo a_n, se tiene que*
$$f(a_n) \sim p_k a_n^k.$$
(Es decir, una función infinitesimal es "equivalente a su desarrollo de MacLaurin", si este es no nulo).

Nótese que la clásica función
$$f(x) = e^{-\frac{1}{x^2}} \text{ si } x \neq 0, \;\; f(0) = 0$$

[2]Para lo cual se requiere que tanto a_n como b_n sean no nulos para n suficientemente grande.

no cumple la condición del lema porque su desarrollo de MacLaurin es 0.

Téngase en cuenta que, casi seguro un límite que aparece como $\frac{0}{0}$ se podrá calcular utilizando la regla de L'Hôpital, salvo que alguna de las sucesiones contenga "términos trigonométricos oscilantes" (i.e. alguna de ellas oscila).

El caso $\frac{\infty}{\infty}$. Se pueden hacer dos cosas: convertirlo en $\frac{0}{0}$ o bien utilizar órdenes de magnitud de infinitos.

Los órdenes de magnitud de infinitos son de la misma naturaleza que los de los infinitésimos (como arriba, $c > 0$):

(6)

$\log(n)$	n^c	c^n	$n!$	n^n

La fórmula de Stirling: Una de las equivalencias de infinitos más importantes que se ha probado es la *fórmula de Stirling*. Es más o menos "mágica", aunque tiene su razón de ser. Esta es una de las pocas que hay que saberse de memoria:

$$n! \sim n^n e^{-n}\sqrt{2\pi n}.$$

Conviene acostumbrarse a calcular límites sustituyendo infinitésimos e infinitos por sus equivalencias (p.ej. "quitando" términos de orden menor, que pueden ser irrelevantes para el cálculo del límite, si es que aparecen sumandos, etc...).

Otras indeterminaciones. Las formas

$$0\cdot\infty,\ \ \infty-\infty,\ \ 1^{\infty},\ \ 0^0,\ \ \infty^0$$

se reducen todas a las ya estudiadas, utilizando las siguientes *igualdades* (donde el logaritmo, es *siempre* el natural):

$$ab = \frac{a}{1/b},\quad a-b = \frac{\frac{1}{b}-\frac{1}{a}}{\frac{1}{ab}},\quad a^b = e^{b\log(a)}$$

y, muchas veces, como $\log(1+x) \equiv x$, si $a_n \to 1$, los límites 1^{∞} pueden ponerse (se supone $b_n \to \infty$):

$$a_n^{b_n} \sim e^{b_n(a_n-1)}$$

con lo que solo hay que calcular el límite de $b_n(a_n-1)$.

Pero *todo lo dicho* hay que explicarlo en los exámenes, no vale "aplicar la fórmula sin más".

1.6. Criterio de Stolz. Llegamos al último (y quizás el único relevante porque es de los únicos que aun no están implementados en los sistemas de computación simbólica). Es el equivalente a la regla de L'Hôpital en sucesiones:

TEOREMA 20 (Criterio de Stolz). *Sean a_n y b_n dos sucesiones de números reales tales que b_n es monótona, no tiene elementos iguales a cero y*

- *O bien b_n es divergente*
- *O bien a_n y b_n convergen ambas a 0.*

Entonces, si existe el límite

$$\lim \frac{a_{n+1}-a_n}{b_{n+1}-b_n},$$

existe también el límite $\lim \frac{a_n}{b_n}$*, y ambos coinciden.*

Este resultado se "utiliza" para calcular límites de expresiones con sumas, por lo general:

$$\lim \frac{1+2+\cdots+n}{n^2}$$

(si uno no sabe que $1+\cdots+n = n(n+1)/2$, se puede hacer):

$$a_n = 1+\cdots+n, \quad b_n = n^2$$

cumplen las condiciones del criterio de Stolz, así que calculamos

$$\lim \frac{a_{n+1}-a_n}{b_{n+1}-b_n} = \frac{n+1}{(n+1)^2-n^2} = \lim \frac{n+1}{2n+1} = \frac{1}{2}.$$

De donde el límite original es $1/2$, también.

2. Series numéricas

EJEMPLO 6. Se sabe que un satélite emite a cada paso por el punto de contacto la mitad de la energía que en el paso anterior. La primera vez emitió $100j$. ¿Cuánta energía emitirá como mucho, en total, al cabo de toda su vida?

Este ejemplo es el clásico problema en que aparece, de manera natural, una sucesión cuyo término general es una suma. Si la energía en el paso k es a_k, entonces la energía acumulada es

$$E_n = a_1 + \cdots + a_n = \sum_{k=1}^{n} a_k$$

y el enunciado pide calcular (o al menos acotar inferiormente)

$$\lim_{n\to\infty} E_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_n$$

DEFINICIÓN 43. Una sucesión cuyo término general es una suma de n términos se llama *serie* y se expresa

$$\sum_{n=1}^{\infty} a_n.$$

Lo cual *no significa que se realice una suma infinita*. Es una manera *breve* de escribir la sucesión

$$E_n = (a_1 + \cdots + a_n).$$

Continuando con el ejemplo, se tiene que $a_1 = 100j$ y se sabe que $a_n = \frac{1}{2}a_{n-1}$, para $n > 1$, de manera que es:

$$E_n = 100 + \frac{100}{2} + \cdots + \frac{100}{2^{n-1}}.$$

(cuidado con el último exponente, es $n-1$, no n). La sucesión E_n se escribe

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{100}{2^{n-1}},$$

que es (como todo el mundo sabe) la sucesión correspondiente a la suma de términos de una progresión geométrica:

$$E_n = \frac{a_1 - a_n r}{1 - r} = \frac{100 - \frac{100}{2^n}}{1 - \frac{1}{2}}$$

Calcular el límite de esta sucesión E_n (que es una *serie*) es sencillo, da: 200.

Esto también se escribe

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{100}{2^{n-1}} = 200,$$

pero solo es una notación: no significa que "la suma infinita dé 200" porque *no sabemos hacer sumas infinitas*. Simplemente indicamos que la sucesión de sumas de n términos de la progresión $\frac{100}{2^{n-1}}$ tiene como límite 200.

Definición 44. Sea S una serie numérica

$$S = \sum_{n=1}^{\infty} a_n$$

de término general a_n. Sea $S_n = \sum_{i=1}^{n} a_n$ la suma $n-$ésima. Se dice que S es

Convergente: Si la sucesión S_n converge.
Divergente: Si la sucesión S_n no converge.

Por otro lado, se dice que es **oscilante** si el término general a_n cambia de signo cada vez. Además, se dice que S es *absolutamente convergente* si la serie $S' = \sum |a_n|$ cuyo término general es el valor absoluto de a_n converge.

Finalmente, si S converge pero no absolutamente, se dice que S es *condicionalmente convergente* (desafortunada nomenclatura).

Si S converge, el límite al que converge se llama *suma de la serie* S.

Se tienen los siguientes resultados, si $S = \sum a_n$ es una serie de término general a_n:

LEMA 16. *Si S converge, entonces $a_n \to 0$.*

DEMOSTRACIÓN. Debería ser capaz de hacerla cualquiera. □

LEMA 17. *Si S es absolutamente convergente, entonces S es convergente.*

DEMOSTRACIÓN. Debería ser sencillo, pues cada vez se suma menos que lo que suma la serie de valores absolutos y si esta converge, la otra debería hacerlo… □

LEMA 18 (Criterio de Leibniz). *Si S es oscilante (es decir, el signo de a_n es el contrario del de a_{n+1}) y $a_n \to 0$, entonces S converge.*

DEMOSTRACIÓN. Esto es lo que debería ser más fácil de probar. □

El problema de las series que no son absolutamente convergentes es que *no se pueden reordenar los términos.* Por ejemplo, la serie

$$S = \sum_{n=1}^{\infty} \frac{(-1)^n}{n},$$

es oscilante. El término general converge a 0, así que S tiene suma finita ¿cuál? Tiene que ver con el logaritmo… Sin embargo, si uno intenta decir "primero sumo los positivos" y "luego sumo los negativos", es decir:

$$S_1 = \sum_{n=1}^{\infty} \frac{1}{2n}, \;\; S_2 = \sum_{n=1}^{\infty} \frac{1}{2n-1},$$

y calcular la suma de S como la suma de S_1 menos la suma de S_2, resulta que está cometiendo un abuso, pues ninguna de las dos series S_1 ni S_2 converge, con lo que la resta de sus sumas no tiene significado

LEMA 19 (La serie armónica). *La serie de término general $\frac{1}{n^\alpha}$:*

- Converge *si $\alpha > 1$.*
- Diverge *si $\alpha \leq 1$.*

Se llama serie armónica.

DEMOSTRACIÓN. No vamos a hacerla en detalle, pero es muy sencillo probarlo utilizando el criterio de Cauchy (sin decir que es "el criterio de Cauchy"): acotar las sumas 2^n-ésimas, unas por arriba y otras por abajo, etc… □

La herramienta fundamental para estudiar el carácter de una serie es el siguiente lemma

LEMA 20. *Sea $S = \sum a_n$ una serie* de términos positivos *y sea $T = \sum b_n$ otra, también de términos positivos.*

- *Si* $b_n \leq a_n$ *y* T *diverge, entonces* S *diverge.*
- *Si* $a_n \leq b_n$ *y* T *converge, entonces* S *converge.*

Se supone que las desigualdades son ciertas a partir de un cierto n_0.

Demostración. Obvio: debería serlo a estas alturas de curso. □

Con este resultado en la mano, se tiene:

Corolario 1 (Comparación con la serie armónica). *Sea* $S = \sum a_n$ *una serie de términos positivos. Entonces*

- *Si* $a_n \leq \frac{1}{n^\alpha}$ *para algún* $\alpha > 1$*, entonces* S *converge.*
- *Si* $a_n \geq \frac{1}{n^\alpha}$ *para algún* $0 \leq \alpha \leq 1$*, entonces* S *diverge.*

Se supone que las desigualdades son ciertas a partir de un cierto n_0.

Demostración. No hay más que utilizar el criterio de la serie mayorante o minorante una vez que se conoce el carácter de la serie armónica. □

Un resultado un poco más elaborado y posiblemente fácil de usar es el *criterio de Pringsheim*:

Lema 21. *Sea* $S = \sum a_n$ *una serie de términos positivos. Supongamos que para cierto* $\alpha > 0$*, existe el límite*

$$\lim_{n\to\infty} a_n n^\alpha = l.$$

Entonces:

- *Si* $l \in \mathbb{R}$ *y* $l \neq 0$*, la serie* S *tiene el mismo carácter que la serie armónica de exponente* α*.*
- *Si* $l = 0$ *y* $\alpha > 1$*, la serie* S *converge.*
- *Si* $l = \infty$ *y* $\alpha \leq 1$*, la serie* S *diverge.*

Damos los otros dos criterios más conocidos:

Lema 22. *Si* $S = \sum a_n$ *es una serie de términos positivos y se tiene que*

$$\lim \frac{a_{n+1}}{a_n} = l \in \mathbb{R},$$

entonces:

- *Si* $l < 1$*, la serie* S *converge.*
- *Si* $l > 1$*, la serie* S *diverge.*
- *Si* $l = 1$*, no se sabe.*

Lema 23. *Si* $S = \sum a_n$ *es una serie de términos positivos y se tiene que*

$$\lim \sqrt[n]{a_n} = l \in \mathbb{R},$$

entonces

- *Si* $l < 1$ *la serie* S *converge.*
- *Si* $l > 1$*, la serie* S *diverge.*
- *Si* $l = 1$*, no se sabe.*

En general, *es mejor acudir a un ordenador que hacer las cuentas a mano*. El siguiente comando en `wolframalpha`

```
sum(n^n/(2n+1)^n,n,1,inf)
```

muestra que la serie

$$\sum_{n=1}^{\infty} \frac{n^n}{(2n+1)^n}$$

converge (por el criterio del cociente) y da una suma aproximada...

DEFINICIÓN 45. Sea $S = \sum a_n$ una serie *convergente* de suma s. Se llama *resto* $n-$*ésimo* a la diferencia:

$$R_n = s - \sum_{k=1}^{n} a_n,$$

que, obviamente, es una sucesión que converge a 0.

Lo que "se hace" en la vida real es *calcular la suma* n*-ésima* y *acotar* el resto $n-$ésimo para saber con "cuánto se acerca uno al resultado de la suma".

DEFINICIÓN 46. La *constante de Euler* (o de Euler-Mascheroni) es el siguiente límite:

$$\lim_{n\to\infty} \left(\log(n) - \sum_{k=1}^{n} \frac{1}{n} \right) \sim 0{,}57721+$$

"La serie armónica diverge como el logaritmo más la constante de Euler", se podría decir.

3. Series de potencias

Fijémonos en la Figura 1. Representa (en negro) la función $\operatorname{sen}(x)$ en el intervalo $[-4, 4]$, aprox. y, en distintos colores, las gráficas de los polinomios de Taylor de la función $\operatorname{sen}(x)$ de grados $1, 3, 5$ y 7.

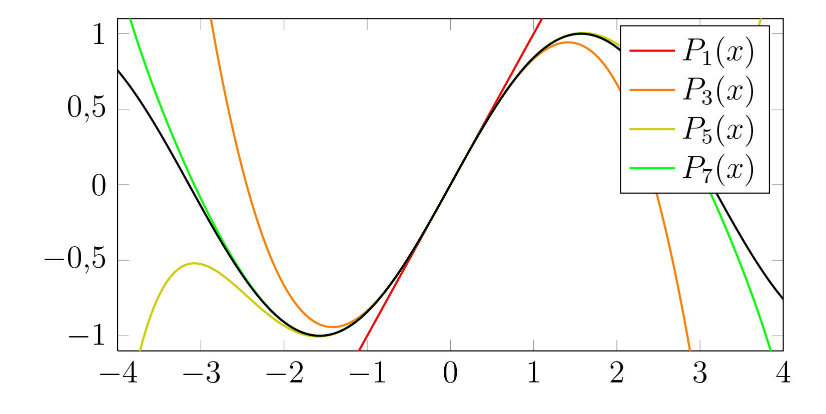


FIGURA 1. Polinomios de Taylor sucesivos

El lector observará que, según aumenta el grado del polinomio, su gráfica “se acerca más” a la de la función seno. El polinomio de grado uno es una recta que en seguida se aleja de la gráfica sinusoidal, el de grado 3 tiene los extremos más cerca de los primeros extremos de la función seno, pero en seguida se “va lejos”, hasta que el de grado 7 está, en la zona visible de la gráfica, muy cerca de la función seno.

El polinomio de Taylor de la función seno cada vez se parece más a la propia función seno, según aumenta su grado. Claro que, según aumenta el grado, se convierte en una *suma cada vez más grande*... Esto nos hace pensar en la posibilidad de escribir una serie no ya numérica, sino *con* x.

Definición 47. Una *sucesión de funciones* es una familia f_n de funciones reales de variable real, una para cada $n \in \mathbb{N}$.

Definición 48. Una *serie de potencias* $S(x)$ es una sucesión de funciones cuyo término general es un polinomio de grado n. Si el término general se escribe como un polinomio en $(x-a)$, donde a es un número real, se dice que $S(x)$ está centrada en a. Se escribe siempre

$$S(x) = \sum_{n=0}^{\infty} a_n (x-a)^n.$$

Por lo general, se toma la suma comenzando en 0.

Está claro que un *polinomio* siempre define una función en todo $\mathbb{R}$, pero ¿una serie de potencias? Podría perfectamente ocurrir que “sustituir la x por un valor” no tuviera sentido. Por ejemplo, si en la serie de potencias

$$S(x) = \sum_{n=0}^{\infty} x^n$$

(centrada en 0) sustituimos la x por $x = 1$, queda

$$S(1) = \sum_{n=0}^{\infty} 1$$

que, *obviamente*, no tiene sentido como número real (la serie que aparece es *divergente*). Pues bien, con las series de potencias ocurre algo especial, que es parte de la clave de su importancia: si la sustitución converge para un cierto número positivo, entonces converge para todos los que tienen valor absoluto menor o igual que ese número.

Teorema 21. *Sea $S(x)$ una serie de potencias centrada en a, es decir:*

$$S(x) = \sum_{n=0}^{\infty} a_n (x-a)^n.$$

Si $S(a+r)$ —es decir, la serie cuyo término general es $a_n(r)^n$— converge para cierto $r > 0$, entonces $S(x)$ converge para todo $x \in [a-r, a+r]$.

De aquí que tenga sentido definir

DEFINICIÓN 49. Dada una serie de potencias $S(x)$ centrada en a, se define el *radio de convergencia* de $S(x)$ como el *sumpremo* de los r tales que $S(a+r)$ converge. Si ese supremo es $+\infty$, entonces se dice que $S(x)$ tiene radio de convergencia infinito.

El radio de convergencia es "sencillo" de calcular:

LEMA 24. *El radio de convergencia de* $S(x) = \sum a_n(x-a)^n$ *es*

$$\lim_{n\to\infty} \sqrt[n]{\frac{1}{|a_n|}}.$$

Lo que indica el radio de convergencia, *a efectos prácticos*, es el intervalo alrededor de a en cual tiene sentido *sumar la serie* $S(x)$.

Pero las series de potencias son útiles sobre todo porque permiten aproximar las funciones que tienen infinitas derivadas, en casos comunes.

DEFINICIÓN 50 (Serie de Taylor). Sea $f : X \to \mathbb{R}$ una función real de variable real que tiene derivadas de todos los órdenes en el punto $a \in \mathbb{R}$. Se denomina *serie de Taylor de f en a* a la serie

$$S(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

que no es más que "el polinomio de Taylor llevado al límite".

Si $a = 0$ (es decir, se hace el desarrollo en 0), hay gente que lo llama *desarrollo de MacLaurin*. Yo no, siempre diré Taylor.

Está claro que para cada x, la serie de Taylor define una serie numérica. Lo que no está tan claro es que esa serie numérica converja. Lo hace en el radio de convergencia:

NOTA 6. La serie de Taylor $S(x)$ de una función $f(x)$ en 0 define otra función en un intervalo $(a-r, a+r)$, donde r es el radio de convergencia de $S(x)$.

El caso es que **puede que la serie de Taylor no defina la misma función que f.**

NOTA 7. Sea $f(x)$ una función con derivadas de todos los órdenes y $S(x) = \sum a_n(x-a)^n$ su desarrollo de Taylor en a. Se define el *resto n-ésimo* de $S(x)$ como

$$R_n(x) = f(x) - \sum_{k=0}^{n} a_n(x-a)^n.$$

Está claro que la serie $S(x)$ tiene como suma $f(x)$ en el punto x si y solo si $R_n(x)$ converge a 0 en ese punto.

Como con el Polinomio de Taylor, si se tiene una cota para la derivada $n-$ésima, se tiene una cota para el error, pues el resto $n-$ésimo de la serie es *lo mismo* que el error cometido por el polinomio de Taylor de grado n.

Recordamos la primera cota del error del polinomio de Tyalor: Teorema 10:

TEOREMA 22 (Cota del error con la derivada). *Si f es una función con derivadas de todos los órdenes en $(a-\varepsilon, a+\varepsilon)$ y $S(x)$ es su serie de Taylor en a y se tiene que $|f^{(n+1)}(x)| < M$ en el intervalo $(a-\varepsilon, a+\varepsilon)$, entonces*

$$|R_n(x)| \leq \frac{1}{(n+1)!} M|x-a|^{n+1},$$

es decir, el error $n-$ésimo es más pequeño que esa cota.

Lo que puede ocurrir es que no se tenga esa cota en el intervalo. Por ejemplo, la función

$$f(x) = e^{-\frac{1}{x^2}} \text{ si } x \neq 0, \ \ f(0) = 0$$

es (como se puede comprobar) derivable infinitas veces en todo $\mathbb{R}$. El valor de la derivada de cualquier orden en 0 es 0, así que la serie de Taylor de $f(x)$ en 0 es

$$S(x) = \sum_{n=0}^{\infty} \frac{0}{n!} x^n = 0.$$

Esta serie converge independientemente del valor de x (siempre suma 0), y por tanto define una función. Como es obvio, la función constantemente igual a 0 *no es* la función $f(x)$ (cuya gráfica se puede ver). ¿Qué ocurre en este caso para que la serie de Taylor sea tan diferente de la función? Que las derivadas de órdenes sucesivos se van haciendo *muy* grandes y los restos $n-$ésimos son también grandes, por tanto.

Así pues, la serie de Taylor de $e^{\frac{-1}{x^2}}$ *es inútil* para aproximar los valores de la función, porque el resto $n-$ésimo es, en este caso, *igual a la función.*

La otra posible pega de los desarrollos de Taylor es que *puede que la serie tenga radio de convergencia nulo.*

EJEMPLO 7. Supongamos que una función $f(x)$ cumple la condición siguiente

$$x^2 f'(x) - f(x) = 1$$

que, por razones obvias, se denomina *ecuación diferencial*, y además, $f(0) = 0$. (Se sabe que esta $f(x)$ existe). Si se calcula su serie de Taylor $S(x)$ en 0, a partir de la ecuación propuesta sale

$$S(x) = \sum_{n=1}^{\infty} (n-1)! x^n$$

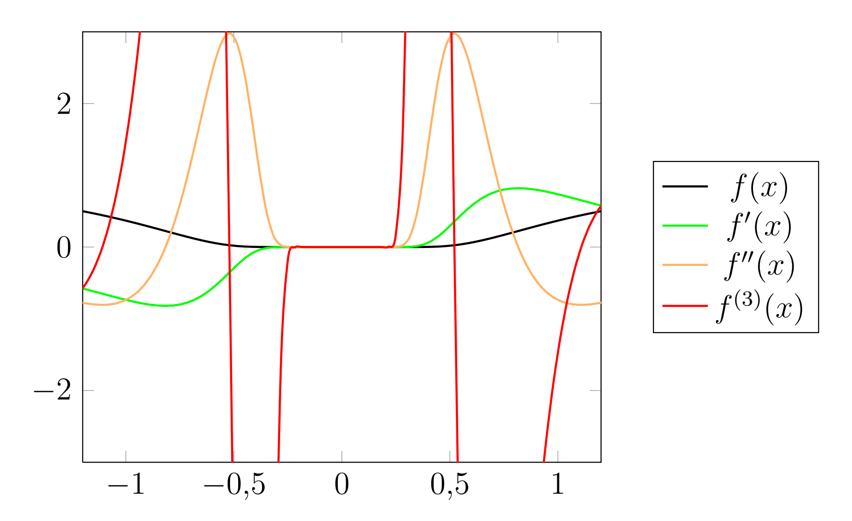


FIGURA 2. La función $e^{\frac{-1}{x^2}}$ y sus derivadas.

y, como todo el mundo sabe,

$$\lim_{n\to\infty} \sqrt[n]{(n-1)!} = \infty,$$

de donde el radio de convergencia es 0. Es decir, si $x_0 > 0$, la serie que se obtiene al sustituir x por x_0 es divergente y por tanto, no se puede decir que la serie de Taylor defina una función.

Así pues, una vez calculada la serie de Taylor de una función $f(x)$, para poder usarla como *aproximación de* $f(x)$ es necesario comprobar que el radio de convergencia es positivo y que el resto $n-$ésimo converge a 0 en todo el intervalo de manera uniforme.

Se sabe algo, pero no mucho. Para la convergencia de la serie a una suma finita en cada punto de un intervalo:

DEFINICIÓN 51. Sea $S(x)$ una serie de potencias

$$S(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$$

y sea $r > 0$ un número positivo. Se dice que $S(x)$ *converge uniformemente en* $[a-r, a+r]$ si para cualquier $\varepsilon > 0$ existe n_0 tal que

$$|R_n(x)| < \varepsilon$$

cuando $n > n_0$ independientemente de $x \in [a-r, a+r]$.

Por ejemplo:

LEMA 25. *Si $S(x)$ es una serie de potencias*

$$S(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$$

y existe $M > 0$ tal que $|a_n| \leq M$ a partir de n_0, entonces $S(x)$ converge uniformemente en cualquier intervalo $[a-r, a+r]$.

Pero, repito, que la serie de Taylor converja uniformemente *no quiere decir* que converja uniformemente a la función. Eso sí, si las derivadas son suficientemente pequeñas, sí que lo hace:

LEMA 26. *Si $f(x)$ es una función que admite todas las derivadas en a y $S(x)$ es su serie de Taylor en a:*

$$S(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

y se sabe que existe $M > 0$ tal que

$$|f^{(n)}(x)| < M^n$$

para todo $n > n_0$ y para todo $x \in [a-r, a+r]$, entonces la serie de Taylor converge a la función $f(x)$ en todos los puntos de $[a-r, a+r]$.

Por ejemplo, esto ocurre con las funciones $\operatorname{sen}(x), \cos(x), \exp(x)$ para $a = 0$ y cualquier r.

Nota: El desarrollo de Taylor de una función en 0 suele denominarse, por motivos históricos, *desarrollo de MacLaurin*, pero intentaré no hacerlo nunca para evitar líos.

4. Derivación e integración de series de potencias

Cuando una serie de potencias define una función en un intervalo (es decir, cuando su radio de convergencia es positivo), ¿cómo se puede calcular la derivada de la función que define? ¿la primitiva? Pues igual que con polinomios, haciéndolo término a término.

TEOREMA 23. *Sea $S(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$ una serie de potencias centrada en a, con radio de convergencia r. Entonces:*

- *La función definida por $S(x)$ en $(a-r, a+r)$ es derivable, y el desarrollo de Taylor de la derivada es*

  $$S_1(x) = \sum_{n=1}^{\infty} na_n(x-a)^n.$$

  Además, este desarrollo converge a la función derivada: $S_1(x) \to S'(x)$ uniformemente en $(a-\varepsilon, a+\varepsilon)$ para $0 < \varepsilon < r$.
- *La función definida por $S(x)$ en $(a-r, a+r)$ es integrable, el desarrollo de Taylor de una primitiva en a es*

  $$S_2(x) = \sum_{n=1}^{\infty} \frac{1}{n+1} a_n(x-a)^{n+1}$$

  y además, este desarrollo converge a la primitiva de $S(x)$ que vale 0 en a:

  $$S_1(x) \to \int_a^x S(x)\,dx$$

  uniformemente en $(a-\varepsilon, a+\varepsilon)$ para $0 < \varepsilon < r$.

La frase "uniformemente en ..." debe entenderse como "sin problemas mayores en todo el intervalo $(a-r, a+r)$", pero está así dicho con precisión.

CAPÍTULO 5

Cálculo [diferencial] en varias variables

Time for a break.

Esto es difícil. Honradamente, solo aspiro a que seáis capaces de saber cómo (y solo en cierto modo “por qué”) se calculan los extremos relativos condicionados o no de funciones de varias variables. Cualquier otra aspiración me parece fuera de lugar. Incluso esta, con el tiempo que hay, me parece poco creíble…

El hecho de utilizar n coordenadas desde el principio no tiene ninguna complejidad. Es mejor acostumbrarse a pasar el trago en seguida y dejarse de historias raras como que “con tres dimensiones tenemos suficiente”. El ejemplo más cercano son los juegos de ordenador, quizás el más notable es “Civilization”: en este juego, no hay un solo valor que determine “quién lo ha hecho mejor”, hay una colección de ellos: crecimiento militar, éxito militar, población, actitud de la población, avance tecnológico, crecimiento económico (ya vamos 6 variables, se ve que tres no eran suficientes para “medir” el “éxito” de una civilización).

Cuando uno calcula los modos de vibración de un sistema físico (por ejemplo un puente o un circuito eléctrico), tiene que tener en cuenta el número de nodos del modelo. Cada nodo define un grado de libertad (más o menos, obviamente), es decir, cada nodo se formaliza como una variable: una dimensión. Hay modelos de elementos finitos con miles (y millones) de nodos… Todo eso son variables. Vectores (o puntos, si se quiere) con millones de coordneadas. ¿y qué?

Etcétera.

1. “Topología” en $\mathbb{R}^n$

El espacio de $n-$uplas de números reales (el producto cartesiano de n copias de $\mathbb{R}$) se denomina $\mathbb{R}^n$. Sus elementos se escriben como una lista de números reales separados por comas y cerrada entre paréntesis:

$$(x_1, x_2, \dots, x_n)$$

si n es dos, se habla de *pares*, si n es tres, se habla de *ternas* o *tríos*, etc…

1.1. Distancia.

Dados dos elementos (puntos) de $\mathbb{R}^n$, la manera más natural de *medir* la relación que hay entre ellos es la *distancia.* En una variable la distancia —por casualidad— coincide con *el valor absoluto de la diferencia.* En varias variables se utiliza la generalización

del Teorema de Pitágoras. Dados dos puntos $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ de $\mathbb{R}^n$, se define la *distancia* entre x e y como:

$$d(x, y) = \|(x_1 - y_1, \dots, x_n - y_n)\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

(las barras verticales paralelas se llaman *norma* de un vector, se habrá estudiado algo relativo a esto en Álgebra, *quizás*.) Es decir, la distancia entre dos puntos es la *longitud* (eso es la norma) del vector que los une.

La noción equivalente a la de *intervalo* en una variable es la de *bola*:

Definición 52. La *bola abierta* de radio r alrededor del punto $a = (a_1, \dots, a_n)$ es el conjunto de puntos $(x_1, \dots, x_n)$ que están a distancia *menor* que r de $(a_1, \dots, a_n)$:

$$B_r(a) = \left\{ x = (x_1, \dots, x_n) \middle| \, d(x, a) < r \right\}.$$

La *bola cerrada* es lo mismo cambiando $<$ por $\leq$:

$$\bar{B}_r(a) = \left\{ x = (x_1, \dots, x_n) \middle| \, d(x, a) \leq r \right\}.$$

Igual que el intervalo, la bola es el conjunto de puntos *alrededor* de a que están más cerca que r. En dimensión 2 (es decir, cuando $n = 2$), la bola abierta es el disco. En dimensión 3, es una esfera sólida, etc. La bola abierta no incluye el "borde" (digamos, la esfera externa), mientras que la bola cerrada sí (en analogía con el intervalo abierto y el cerrado). Se llama *bola punteada* a la bola abierta *sin el centro* (en la definición, sin el punto a), la denotaremos $\dot{B}_r(a)$.

Un punto está "pegado" a un conjunto si cualquier bola alrededor de ese punto corta al conjunto:

Definición 53. Un punto $a \in \mathbb{R}^n$ es *adherente* a un conjunto $X \subset \mathbb{R}^n$ si cualquier bola centrada en a tiene intersección con X.

Está claro que los puntos de X son adherentes a X y esto, si bien es intersante, no refleja la naturaleza de los puntos "muy cercanos" pero no necesariamente dentro de X:

Definición 54. Un punto $a \in \mathbb{R}^n$ es *de acumulación* de un conjunto $X \in \mathbb{R}^n$ si cualquier bola *punteada* centrada en a corta a X.

Por ejemplo, si $X = \left\{ \frac{1}{n} \middle| \, n \in \mathbb{N} \right\}$ (los puntos de la sucesión $1/n$), entonces los puntos $1, \frac{1}{2}, \dots$ son todos obviamente *adherentes* a X, mientras que el 0 es *además* un punto de acumulación (que en este caso no pertenece a X, pero podría hacerlo).

El concepto de conjunto abierto y cerrado puede ser útil, aunque aquí nos servirá más que nada para ser precisos

Definición 55. Un conjunto $X \in \mathbb{R}^n$ se dice *abierto* si dado cualquier $a \in X$ existe una bola centrada en a y totalmente contenida en X. Así, un abierto es un conjunto "relleno por todas partes".

Un conjunto $X \in \mathbb{R}^n$ se dice *cerrado* si cualquier punto de acumulación de X pertenece a X. Es decir, un cerrado es un conjunto que "contiene su borde".

Diremos que X es un *entorno* de un punto $a \in \mathbb{R}^n$ si hay una bola centrada en a contenida en X. (Por tanto, todas las bolas de radio más pequeño también lo están). Un entorno de un punto es un conjunto que, alrededor de ese punto está "relleno".

DEFINICIÓN 56. Diremos que una función $f : X \to \mathbb{R}$ es un *infinitésimo en* $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ —o que tiene límite 0 en a— si

- X es un entorno de a y
- Para cualquier $\varepsilon > 0$, existe una distancia $\delta > 0$ tal que si $d(x, a) < \delta$ entonces $|f(x)| < \varepsilon$.

Nótese que para definir infinitésimo se impone que $f(a) = 0$ (pues si no, no se cumple la segunda condición). Hablaremos de *infinitésimo* sin más cuando a sea $(0, \dots, 0)$.

Exactamente como en el caso de una variable, la definición de infinitésimo "signfica" que "la función se hace muy pequeña cerca de a" (es decir, se hace "arbitrariamente pequeña" se acerque uno como se acerque a 0). Como se ve, se utiliza la única medida de "distancia" posible que hay en $\mathbb{R}^n$, que es la dada arriba por la norma. *No se puede comprobar si algo converge a cero mirando en una u otra dirección.*

EJEMPLO 8. Considérese la función en $\mathbb{R}^2$

$$f(x, y) = \begin{cases} 0 \text{ si } xy = 0 \\ \frac{1}{xy} \text{ si } xy \neq 0 \end{cases}$$

Es obvio que $\lim_{x\to 0} f(x, 0) = 0$ y que $\lim_{y\to 0} f(0, y) = 0$. Pero también debería ser obvio que f *no es un infinitésimo* en $a = (0, 0)$.

DEFINICIÓN 57. Sea $f : X \to \mathbb{R}$ una función definida en un subconjunto de $\mathbb{R}^n$. Se dice que f es *continua en* $a \in \mathbb{X}$ si existe un infinitésimo en a, digamos $\mathrm{d}(x)$ y una distancia r tal que

$$f(x) = f(a) + \mathrm{d}(x)$$

para todo $x \in X \cap B_r(a)$ (Usamos esta intersección porque X no tiene por qué ser un entorno de a, aunque lo será siempre en las aplicaciones).

Así pues, una función continua en un punto a es una función que, cerca de a vale "lo que vale en a más un infinitésimo". De este modo, una función es continua si se *aproxima a* $f(a)$ "cerca" de a.

Una noción menos relevante es la de *límite* de una función en un punto.

DEFINICIÓN 58. Se dice que l es el límite de f en a (con f y a como arriba) si existe un infinitésimo en a, digamos $\mathrm{d}(x)$ tal que

$$f(x) = l + \mathrm{d}(x)$$

para x en una cierta bola *punteada* $x \in \dot{B}_r(a)$.

Desde luego, una función es continua en a si tiene límite en a y este coincide con $f(a)$. Vamos a estudiar muy pocas funciones con límite pero no continuas.

EJEMPLO 9. ¿Qué pasa con la función $f(x,y) = \text{sen}(\frac{1}{xy})$ (poniendo $f(x,y) = 0$ si $xy = 0$) en $a = (0,0)$? Lo mismo que con la función de la figura 7 (que no es continua en $(0,0)$). Ni siquiera se puede definir un "límite" en ese punto.

DEFINICIÓN 59. Se dice que $f : X \to \mathbb{R}$ es continua en X si es continua en todos sus puntos.

Antes de enunciar los teoremas equivalentes a los de una variable, necesitamos la noción de *conjunto compacto* (una de las más fructíferas):

DEFINICIÓN 60. Un conjunto $X \subset \mathbb{R}^n$ se dice *compacto* si es cerrado y acotado. Es decir, si es cerrado y existe $R > 0$ tal que $X \subset B_R(0)$ (si cabe dentro de una bola lo suficientemente grande).

1.2. Los teoremas de la continuidad. El primer resultado importante es:

TEOREMA 24. *La imagen de un compacto por una aplicación continua es un compacto.*

De donde se deduce que:

COROLARIO 2 (Weierstrass en n variables). *Si $f : X \to \mathbb{R}$ es una aplicación continua de un subconjunto de $\mathbb{R}^n$ en $\mathbb{R}$ y X es compacto, entonces existen $a, b \in \mathbb{X}$ tales que $f(a)$ es el máximo de $f(X)$ y $f(b)$ es el mínimo de $f(X)$. [Recuérdese que $f(X)$ es el conjunto de valores que toma f, el* recorrido*]*

Por otro lado, es evidente del Teorema de Bolzano que:

TEOREMA 25. *Si $f : X \to \mathbb{R}$ es continua y existen $a, b \in X$ tales que $f(a) < f(b)$ y el segmento que une a y b está incluido en X, entonces existe $c \in X$ tal que $f(c) = 0$.*

De hecho, el resultado que se puede obtener es mucho más fuerte, pero lo dejamos para más adelante (cuando definamos *curva*).

1.3. Funciones con valores en $\mathbb{R}^m$. Igual que podemos definir funciones $f : \mathbb{R}^n \to \mathbb{R}$, se puede pensar en funciones que comienzan en $\mathbb{R}^n$ y terminan en $\mathbb{R}^m$: lo que clásicamente se conoce como *funciones vectoriales*. Las denotaremos utilizando el vector de sus componentes: $f : \mathbb{R}^n \to \mathbb{R}^m$ no es más que una colección de m funciones $f_i : \mathbb{R}^n \to \mathbb{R}$.

Ejemplo 10. En la superficie delimitada por el mapa de España (es decir, en un conjunto de $\mathbb{R}^2$), se puede poner en cada punto el vector dado por la *velocidad del viento* en un instante. En cada punto (x, y) del mapa se tendrá un vector (en este caso también de 2 componentes) $(v_1(x, y), v_2(x, y))$. Esto es una aplicación de $\mathbb{R}^2$ en $\mathbb{R}^2$.

Ejemplo 11. Con la misma idea de antes, en lugar de solo el vector velocidad, se puede asignar a cada punto del plano la velocidad del viento, la temperatura y la presión atmosférica. Esto da una aplicación

$$f : \mathbb{R}^2 \to \mathbb{R}^4$$

cuyas componentes son, por ejemplo, $(v_1(x, y), v_2(x, y), t(x, y), p(x, y))$.

Ejemplo 12. Dado un intervalo $[r, s] \subset \mathbb{R}$, se define una *curva parametrizada por* $[r, s]$ como una aplicación $\gamma : [r, s] \to \mathbb{R}^n$ (para un n determinado) que cumple cierta condición (que enunciaremos en un momento). La aplicación γ define una asignación $t \mapsto (\gamma_1(t), \dots, \gamma_n(t))$, que a cada número t (piénsese en t como "la variable temporal tiempo") hace corresponder un punto del espacio $\mathbb{R}^n$, es decir, a cada número le corresponde una $n-$upla de números (los $\gamma_i(t)$ para $i = 1, \dots, n$). Se dice que γ es una *curva* si cada una de las componentes $\gamma_i(t)$ es una función continua.

Es obvio que las funciones *vectoriales* no son más que colecciones de funciones que toman valores en $\mathbb{R}$. Así, es lógico que se defina

Definición 61. Una función vectorial $f : \mathbb{R}^n \to \mathbb{R}^m$ es *continua* si cada una de sus componentes es continua.

En el ejemplo 11, es lógico pensar que tanto las componentes de la velocidad del viento como la temperatura en cada punto y la presión son funciones continuas (pero ¿por qué es lógico esto?). Es lógico pedirle a una curva que sea continua (porque uno espera que una *curva* se pueda *trazar*, si no sería un "salto", no una curva).

Definición 62. Un conjunto $X \subset \mathbb{R}^n$ se dice *conexo* si dados dos puntos $a, b \in X$, existe siempre una curva que comienza en a y termina en b.

Un conjunto es, por tanto, conexo si dos puntos que le pertenecen *se pueden conectar* siempre (de ahí el nombre).

Lema 27. *Sean $f : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^m \to \mathbb{R}^l$ dos funciones continuas. La función compuesta $g \circ f : \mathbb{R}^n \to \mathbb{R}^l$ es continua.*

Demostración. Obvio (no hay más que usar la condición de infinitésimo para cada coordenada, etc…). □

El Teorema de Bolzano bien extendido a $\mathbb{R}^n$ es:

TEOREMA 26. *Sea $X \subset \mathbb{R}^n$ un subconjunto conexo y $f : X \to \mathbb{R}$ una función continua en X. Si existen $a, b \in X$ tales que $f(a) > 0$ y $f(b) < 0$, entonces existe al menos un $c \in X$ tal que $f(c) = 0$.*

DEMOSTRACIÓN. Debería ser sencillita... No hay más que "trasladar el problema a $\mathbb{R}$" mediante la curva que une a y b y utilizar ahí el teorema de Bolzano. □

Como se ve, toda la idea de continuidad se refiere a funciones para las que tiene sentido la noción de *aproximación* (una función es continua en un punto si su valor *se aproxima* al valor en el punto cuando la variable se aproxima al punto). De hecho, todo el Análisis es una teoría de la aproximación (y cada condición que se estudia es una "mejora" de las propiedades aproximativas). La continuidad es la noción más simple de aproximación. La siguiente que estudiaremos es la *diferenciabilidad* que se corresponde con la derivabilidad en una variable.

2. Diferencial de una función

En una variable, una aplicación $f : \mathbb{R} \to \mathbb{R}$ es *derivable* en x_0 si existen un número λ y un infinitésimo $\mathrm{d}(h)$

$$f(x_0 + h) = f(x_0) + \lambda h + \mathrm{d}(h)h \tag{7}$$

para h "pequeño". Lo escrito arriba es obviamente equivalente a que exista el límite

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

y que sea igual a λ, que es la "definición de derivada como pendiente".

Pensando en *aproximaciones*, lo que tenemos es que f es derivable en x_0 si f es, cerca de x_0 el valor $f(x_0)$ más una función $h \mapsto \lambda h$ más un infinitésimo *multiplicado por h*, que mide "cuánto de mal aproxima la recta tangente a la función en x_0". Esta es la noción que vamos a traspasar a $\mathbb{R}^n$, la noción de *aproximación*.

DEFINICIÓN 63. Sea $f : X \to \mathbb{R}$ una función definida en $X \subset \mathbb{R}^n$ y sea $a = (a_1, \dots, a_n) \in X$ un elemento de X. Se supone que X *es un entorno de a*. Se dice que f es *diferenciable* en a si existen una aplicación lineal $A : \mathbb{R}^n \to \mathbb{R}$ y un infinitésimo $\mathrm{d}(h)$ tales que

$$f(a + h) = f(a) + A(h) + \|h\|\mathrm{d}(h),$$

donde $\|h\|$ es, como siempre, la norma del vector h.

Es decir, una función de n variables es diferenciable si *se parece mucho* a una aplicación lineal. "Mucho" significa que la diferencia entre la función y la aplicación lineal es infinitamente más pequeña que "la longitud" del vector h. Como se vio al hablar de infinitésimos en una variable (Definición 20), la comparación entre infinitésimos se hace dividiendo, y como no sabemos dividir por vectores, se utiliza la norma (el "módulo") para medir el tamaño de un vector y a la hora de dividir.

Es decir, una función diferenciable en a se podrá escribir

$$f(a+h) = f(a) + (\lambda_1 h_1 + \cdots + \lambda_n h_n) + \mathrm{d}(h)\|h\|,$$

de manera que $f(a+h) - f(a)$ será "casi igual" a una combinación lineal de las componentes de h, pues estas son las únicas aplicaciones lineales que conocemos (¿o no?).

NOTA 8. Recuérdese eso: si $A : \mathbb{R}^n \to \mathbb{R}$ es una aplicación lineal, tomando la base estándar, $e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$, se tendrán valores $\lambda_1 = A(e_1), \dots, \lambda_n = A(e_n)$ de manera que si $h = (h_1, \dots, h_n)$ es un vector cualquiera de $\mathbb{R}^n$, entonces

$$A(h) = \begin{pmatrix} \lambda_1 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \lambda_1 h_1 + \cdots + \lambda_n h_n.$$

De ahí la expresión del párrafo anterior.

Para aplicaciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ la definición es "exactamente la misma":

DEFINICIÓN 64. Una función $f : X \to \mathbb{R}^m$ de un subconjunto $X \subset \mathbb{R}^n$ es diferenciable en x_0 (se supone, como antes, que X es un entorno de x_0) si cada una de sus componentes es diferenciable.

De hecho, (no lo vamos a demostrar), se tiene el siguiente resultado:

TEOREMA 27. *Una función vectorial $f : X \to \mathbb{R}^m$ es diferenciable en x_0 si y solo si existen una aplicación lineal $A : \mathbb{R}^n \to \mathbb{R}^m$ y un infinitésimo* d *tales que*

$$f(x_0 + h) = f(x_0) + A(h) + \|h\|\mathrm{d}(h),$$

"cerca" de x_0. Es decir, si f es "prácticamente" una aplicación lineal cerca de x_0.

De la misma manera que antes, la aplicación lineal A tendrá una cierta matriz en las bases estándar, pero nos estamos adelantando

Notación: de ahora en adelante, si $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en un punto $a \in X$ (según la definición de arriba), la aplicación lineal de la definición se llamará la *diferencial de f en a* y se denotará $D_a(f)$ o, más sencillamente, $D_a f$.

2.1. Funciones derivables obvias. Antes de seguir, enunciamos los resultados obvios de diferenciabilidad de funciones, para poder poner ejemplos.

TEOREMA 28. *Sean $f, g : X \subset \mathbb{R}^n \to \mathbb{R}$ dos aplicaciones y $a \in X$ un punto de X del cual este es un entorno. Entonces:*

- *Si f es constante en un entorno de a, f es diferenciable en a y su diferencial es la aplicación lineal nula:* 0.

- *Si f y g son diferenciables en a, entonces $f \pm g$ y fg son diferenciables en a. Si además $g(a) \neq 0$, entonces f/g también es diferenciable en a. Además:*

$$D_a(f+g) = D_a(f) + D_a(g)$$
$$D_a(fg) = g(a)D_a(f) + f(a)D_a(g)$$
$$D_a(f/g) = \frac{g(a)D_a(f) - f(a)D_a(g)}{g(a)^2}$$

- *Las aplicaciones "coordenadas" $f(x_1, \dots, x_n) = x_i$ son diferenciables en a y*

$$D_a(x_i)\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = h_i$$

  (o sea, la diferencial es la proyección en la $i-$ésima coordenada).

Con esto tenemos, por ejemplo, que cualquier expresión racional en las coordenadas es diferenciable en los puntos en que el denominador no se anula.

3. Derivadas parciales - matriz jacobiana

Fijemos una aplicación $f : X \subset \mathbb{R}^n \to \mathbb{R}$ y un punto $a \in X$ tal que X es un entorno de a (de ahora en adelante utilizaremos la nomenclatura clásica y diremos que a es un *punto interior de* X.) Supongamos que f es diferenciable en a. Como se ha visto, esto quiere decir que existe una aplicación lineal $D_a f$ que *aproxima muy bien a* f, es decir, tal que existe un infinitésimo $\mathrm{d}(h)$ tal que

$$f(a+h) = f(a) + (D_a f)(h) + \|h\|\mathrm{d}(h).$$

Pero en la vida real *utilizamos coordenadas para todo*. En las bases estándar, el vector h (el "incremento de la variable") tendrá coordenadas $h = (h_1, \dots, h_n)$ y su norma será $\sqrt{h_1^2 + \dots + h_n^2}$. En estas coordenadas, la aplicación lineal tendrá una expresión

$$D_a(f) \simeq \begin{pmatrix} \lambda_1 & \dots & \lambda_n \end{pmatrix}$$

de manera que, como dijimos arriba, quedará

$$f(a+h) = f(a) + \begin{pmatrix} \lambda_1 & \dots & \lambda_n \end{pmatrix}\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} + \sqrt{h_1^2 + \dots + h_n^2}\mathrm{d}(h).$$

Ahora bien. Supongamos que, en lugar de utilizar un vector genérico, tomamos uno que es proporcional a un elemento de la base estándar: p.ej. el primero, $h = (h_1, 0, \dots, 0)$. Queda

$$f(a_1 + h_1, a_2, \dots, a_n) = f(a_1, \dots, a_n) + \lambda_1 h_1 + h_1\eta(h_1)$$

donde η es un infinitésimo (importa poco cómo es, pero es infinitésimo). Es decir, *si se dejan constantes todas las coordenadas del punto* $(a_1, \dots, a_n)$ *menos la primera*, el hecho de que f sea diferenciable quiere decir que la aplicación

$$f_1 : \mathbb{R} \to \mathbb{R}$$

dada por $f_1(x) = f(x, a_2, \dots, a_n)$ es una función de una variable *derivable en* a_1.

El mismo razonamiento sirve para el resto de coordenadas:

TEOREMA 29. *Si* $f : X \subset \mathbb{R}^n \to \mathbb{R}$ *es diferenciable en un punto* $a \in X$ *interior a* X*, entonces cada una de las funciones*

$$f_i : \mathbb{R} \to \mathbb{R}$$

de una variable que consisten en "dejar fijas todas las coordenadas menos la i" son derivables en a_i.

Las derivadas ${f_i}'(a_i)$ *se llaman* derivadas parciales de f *(con respecto a la coordenada* i*, respectivamente) y se denotan:*

$$\left.\frac{\partial f}{\partial x_i}\right|_a \text{ o bien } \frac{\partial f}{\partial x_i}(a).$$

Cuando se entiendan las derivadas parciales como funciones en X *(pues es lo que son, al fin y al cabo, para cada elemento interior a* X *en el que existen, toman un valor), se denotarán sin hacer referencia al punto:*

$$\frac{\partial f}{\partial x_i}.$$

Del teorema recién enunciado y de toda la discusión anterior se deduce, que:

TEOREMA 30. *La matriz de* $D_a(f)$ *en la base estándar es la matriz de las derivadas parciales de* f *en* a*:*

$$D_a(f) \simeq \left(\left.\frac{\partial f}{\partial x_1}\right|_a \quad \dots \quad \left.\frac{\partial f}{\partial x_n}\right|_a \right).$$

Esta matriz se denomina matriz jacobiana *de* f.

Por la misma razón, dada la definición de función vectorial diferenciable (Definición 64), se tiene que

TEOREMA 31. *La matriz de la diferencial de una función vectorial* $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ *en un punto* a en las bases estándar *es la matriz formada por las derivadas parciales de cada coordenada de* f*:*

$$D_a(f) = \begin{pmatrix} \left.\frac{\partial f_1}{\partial x_1}\right|_a & \dots & \left.\frac{\partial f_1}{\partial x_n}\right|_a \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial f_m}{\partial x_1}\right|_a & \dots & \left.\frac{\partial f_m}{\partial x_n}\right|_a \end{pmatrix}.$$

Esta matriz se denomina matriz jacobiana de f en a. *Se denotará, a veces,* $J_a(f)$.

EJEMPLO 13. Comencemos con un ejemplo simple: una aplicación afín tiene que tener una diferencial "esencialmente igual" a ella. Sea $f(x,y,z) = 3 - 2x + 4y - 7z$. ¿Es diferenciable? Sabemos, por el Teorema 28 que sí, que lo es en todo punto. Sea $p = (x,y,z)$ un punto cualquiera. ¿Cuál es la matriz jacobiana de f en p? Hemos de calcular las parciales. Para ello, *se dejan constantes todas las variables menos la que se deriva*:

$$\left.\frac{\partial(3-2x+4y-7z)}{\partial x}\right|_p = -2, \quad \left.\frac{\partial(3-2x+4y-7z)}{\partial y}\right|_p = 4,$$
$$\left.\frac{\partial(3-2x+4y-7z)}{\partial z}\right|_p = -7,$$

es decir, que por la definición de diferenciabilidad, queda que

$$f(p+h) = f(p) + \begin{pmatrix} -2 & 4 & -7 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} + \|h\|\mathrm{d}(h)$$

para un infinitésimo $\mathrm{d}(h)$. Si $p = (0,0,0)$, por ejemplo, lo que queda es (sabemos que $f(0,0,0) = 0$:

$$f(h_1,h_2,h_3) = 3 - 2h_1 + 4h_2 - 7h_3 + \|h\|\mathrm{d}(h)$$

pero, también sabemos que $f(x,y,z) = 3 - 2x + 4y - 7z$, así que en este caso concreto *la diferencial vale lo mismo que la parte lineal de la función* y la aproximación es obviamente exacta, $\mathrm{d}(h) = 0$.

EJEMPLO 14. Un ejemplo algo más elaborado. Consideremos la aplicación diferencible $f(x,y) = 1 - x^2 - y^2$ (cuya gráfica es un paraboloide de revolución "hacia abajo"). Está claro que la función $f(x,y)$ tiene un máximo en $(0,0)$ —¿está esto claro? Calculemos su diferencial. Como siempre se hace, utilizamos las parciales. En un punto (x,y) estas son:

$$\frac{\partial f}{\partial x}(x,y) = -2x, \quad \frac{\partial f}{\partial y}(x,y) = -2y,$$

de modo que $\frac{\partial f}{\partial x}(0,0) = 0$ y $\frac{\partial f}{\partial y}(0,0) = 0$. Las derivadas parciales son ambas 0 en el origen y por tanto la aplicación diferencial es la nula $D_{(0,0)}f \simeq (0\ \ 0)$.

Debería ser obvio que las parciales tienen que ser 0 en el origen. Si se considera la función $f_1(x) = f(x,0) = 1-x^2$, esta es una función de una variable que tiene un máximo en $x = 0$, así que su derivada en 0 tiene que ser nula. Lo mismo ocurre si se considera $f_2(y) = f(0,y) = 1 - y^2$. Por eso las parciales son nulas, porque el origen es un máximo de f al igualar una de las coordenadas a 0.

En otros puntos, está claro que alguna de las derivadas parciales no es nula, pues $\frac{\partial f}{\partial x} = -2x$ y $\frac{\partial f}{\partial y} = -2y$.

EJEMPLO 15. Una función vectorial: sea $f : \mathbb{R}^3 \to \mathbb{R}^4$ la función que a cada punto del espacio le asigna el vector $(x^3, 2 - xy, z + x - y, zxy)$. Sabemos que es diferenciable por el Teorema 28. ¿Cuál es su matriz jacobiana en un punto (x, y, z)? Fácil:

$$\begin{pmatrix} 3x^2 & 0 & 0 \\ -y & -x & 0 \\ 1 & -1 & 1 \\ zy & zx & xy \end{pmatrix}$$

…¿por qué? ¿por qué tiene tres columnas y cuatro filas?

4. Regla de la cadena

Uno de los resultados básicos es la regla de la cadena:

TEOREMA 32 (Regla de la Cadena). *Sean $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ y $g : Y \subset \mathbb{R}^m \to \mathbb{R}^p$ dos aplicaciones vectoriales. Supongamos que $a \in X$ es un punto interior de X y que $b = f(a) \in Y$ es un punto interior de Y. Si f es diferenciable en a y g es diferenciable en $b = f(a)$, entonces $g \circ f$ es diferenciable en a y su diferencial es la composición de $D_a f$ con $D_{f(a)} g$, es decir:*

$$D_a(g \circ f) = D_{f(a)} g \circ D_a f.$$

DEMOSTRACIÓN. Esto es prácticamente evidente utilizando la definición de diferencial como aplicación lineal que aproxima. Se sabe que

$$f(a + h) = f(a) + D_a f(h) + \|h\| \mathrm{d}(h)$$

y que

$$g(f(a) + \overline{h}) = g(f(a)) + D_{f(a)} g(\overline{h}) + \|\overline{h}\| \mathrm{d}_2(\overline{h}) \tag{8}$$

para ciertos infinitésimos (vectoriales pero es irrelevante) d_1 y d_2. Por consiguiente:

$$g(f(a + h)) = g(f(a) + D_a f(h) + \|h\| \mathrm{d}_1(h))$$

que es igual, por ser g diferenciable, a

$$g(f(a + h)) = g(f(a)) + (D_{f(a)} g)(D_a f(h) + \|h\| \mathrm{d}_1(h)) + \\ + \|D_a f(h) + \|h\| \delta(h)\| \mathrm{d}_2(D_a f(h) + \|h\| \mathrm{d}_1(h)).$$

Es relativamente sencillo comprobar que $\|D_a f(h) + \|h\| \mathrm{d}_1(h)\|$ es un infinitésimo del tipo $\|h\| \mathrm{d}_3(h)$ y que $\mathrm{d}_2(D_a f(h) + \|h\| \mathrm{d}_1(h))$ es un infinitésimo en h, por lo que:

$$g(f(a + h)) = g(f(a)) + (D_{f(a)} g)(D_a f(h) + \|h\| \mathrm{d}_1(h)) + \|h\| \mathrm{d}_4(h)$$

para cierto inifnitésimo $\mathrm{d}_4(h)$. Por otro lado, $(D_{f(a)}g)$ es un aplicación lineal, así que

$$(D_{f(a)}g)(D_a f(h) + \|h\|\mathrm{d}_1(h)) = (D_{f(a)}g)(D_a f(h)) + \|h\|(D_{f(a)}g)(\mathrm{d}_1(h))$$

y la función $(D_{f(a)}g)(\mathrm{d}_1(h))$ es (se comprueba fácilmente por ser lineal $D_{f(a)}g$) un infinitésimo en h. De donde la expresión inicial (8) queda

$$g(f(a+h)) = g(f(a)) + (D_{f(a)}g)(D_a f(h)) + \|h\|\mathrm{d}_5(h)$$

para cierto infinitésimo $\mathrm{d}_5(h)$. La composición de aplicaciones lineales $D_{f(a)}g$ y $D_a f$ es lineal y, por tanto, $g \circ f$ es diferenciable en a.

En esta demostración se han introducido "muchos" infinitésimos con nombres d_i. En realidad, el "Físico" que llevamos dentro los llamaría a todos $\mathrm{d}(h)$ y "a correr": lo único relevante es que son *infinitésimos*, no sus nombres. □

Puede parecer todo "muy técnico", pero de este resultado se deduce:

TEOREMA 33. *Si* $f : \mathbb{R}^n \to \mathbb{R}$ *es diferenciable en* a*, entonces*

- *La exponencial* $\exp(f(x))$ *es diferenciable en* a*,*
- *Las funciones trigonométricas de* f *son diferenciables en* a *(donde estén definidas),*
- *Si* $f(x) > 0$ *entonces* $\log(f(x))$ *es diferenciable en* x*,*
- *Si* $f(x) > 0$ *y* $c \in \mathbb{R} > 0$ *entonces* $f(x)^c$ *es diferenciable en* x*,*
- *La función* $f(x)^{g(x)}$ *es diferenciable siempre que esté definida*[1]*,*
- *Etc...*

Es decir, "todo" lo que se puede escribir con las operaciones normales con la funciones coordenadas y allá donde tenga sentido, "va a ser" diferenciable.

COROLARIO 3. *La regla de la cadena es muy sencilla de utilizar con las matrices jacobianas. Sean* f *y* g *como en el Teorema 32. Entonces la matriz jacobiana de* $g \circ f$ *en un punto* a *es el producto de la matriz jacobiana de* g *en* $f(a)$ *por la matriz jacobiana de* f *en* a*. Es decir,*

$$J_a(g \circ f) = J_{f(a)}g \cdot J_a(f),$$

exactamente igual *que en una variable (pues una variable es un caso particular de varias).*

EJEMPLO 16. La función $f : \mathbb{R}^3 \to \mathbb{R}$ dada por

$$f(x, y, z) = \operatorname{sen}(-x + 2y - 3z)$$

es la composición de $h(x, y, z) = -x + 2y - 3z$ con $g(t) = \operatorname{sen}(t)$. Sabemos que la matriz jacobiana de h es $(-1\ 2\ -3)$ mientras que

[1]Estamos abusando de la notación muchísimo.

la matriz jacobiana de g en un punto t es $(\cos(t))$. Así que la matriz jacobiana de f en (x, y, z) es:

$$J_{(x,y,z)}f = J_{-x+2y-3z}(g) \cdot J_{(x,y,z)}(h) = \begin{pmatrix}\cos(-x+2y-3z)\end{pmatrix} \begin{pmatrix}-1 & 2 & -3\end{pmatrix} =$$
$$= \begin{pmatrix}-\cos(-x+2y-3z) & 2\cos(-x+2y-3z) & -3\cos(-x+2y-3z)\end{pmatrix},$$

que, claramente, tiene 1 fila y 3 columnas, pues la función f va de $\mathbb{R}^3$ a $\mathbb{R}$.

Si tratamos de calcular esa matriz jacobiana utilizando las parciales, es igual de sencillo. Con respeto a x, se tiene (hay que dejar y y z constantes):

$$\left.\frac{\partial(\operatorname{sen}(-x+2y-3z))}{\partial x}\right|_{(x,y,z)} = -\cos(-x+2y-3z).$$

Con respecto a y:

$$\left.\frac{\partial(\operatorname{sen}(-x+2y-3z))}{\partial y}\right|_{(x,y,z)} = 2\cos(-x+2y-3z).$$

Y, finalmente,

$$\left.\frac{\partial(\operatorname{sen}(-x+2y-3z))}{\partial y}\right|_{(x,y,z)} = -3\cos(-x+2y-3z).$$

Por tanto,

$$J_{(x,y,z)}f =$$
$$= \begin{pmatrix}-\cos(-x+2y-3z) & 2\cos(-x+2y-3z) & -3\cos(-x+2y-3z)\end{pmatrix},$$

como arriba (… *claro*…).

5. ¿Plano tangente? ¿¿curvas de nivel??

Vamos a hacer algo de geometría y eso…

Por un momento nos restringimos a $\mathbb{R}^2$. Si $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ es una función real en un subconjunto del plano, puede *dibujarse* (no es necesariamente *sencillo*). Es relativamente sencillo en WolframAlpha. Sea $f(x, y) = x^2 + y^2 - \operatorname{sen}(xy)$, si se introduce

```
plot(x^2+y^2-sin(x y),x=[-2,2],y=[-2,2])
```

se obtiene un hermoso dibujo tridimensional donde la superficie pintada representa el valor (en la coordenada z) de la función f en el punto (x, y). El cuadrado base es $[-2, 2] \times [-2, 2]$, eso es lo que significan el segundo y tercer parámetros del comando.

Además, aparece otro dibujo "titulado" *Contour plot*. Representa, como veremos más adelante, las *curvas de nivel*: cada curva indica una zona en la que la función toma el mismo valor. En el ejemplo, son una especie de óvalos inclinados cuarenta y cinco grados (por el xy del seno).

Olvidemos los gráficos anteriores y centrémonos en una función más sencilla. Ya hemos estudiado el paraboloide de revolución en el Ejemplo

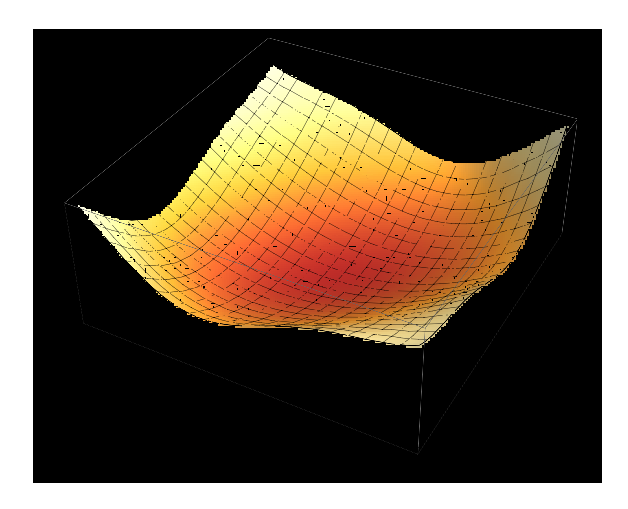


FIGURA 1. Gráfica de $x^2 + y^2 - \text{sen}(xy)$ por WolframAlpha.

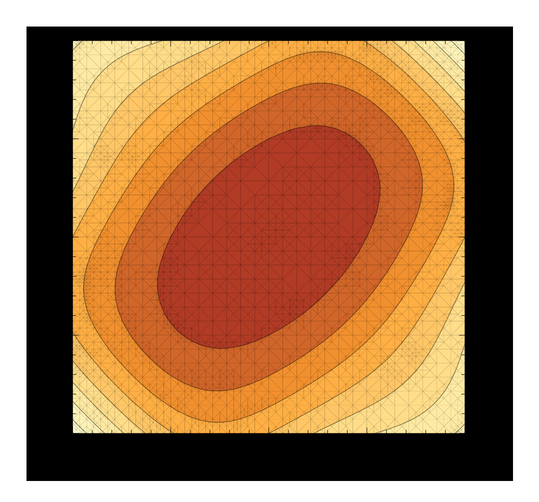


FIGURA 2. Curvas de nivel de $x^2 + y^2 - \text{sen}(xy)$ por WolframAlpha.

14. Volvamos a él. Su ecuación era $f(x, y) = 1 - x^2 - y^2$. Es decir, su gráfica viene dada por la ecuación $z = 1 - x^2 - y^2$.

Si calculamos su diferencial en un punto (a, b), sale

$$J_{(a,b)}(f) = -2a - 2b.$$

Es decir, que *cerca del punto* (a, b), la función $f(x, y)$ se puede representar como

$$f(a + h_1, b + h_2) = f(a, b) + \begin{pmatrix} -2a & -2b \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \|h\| \text{d}(h)$$

para cierto infinitésimo $\text{d}(h)$. Si ahora nos olvidamos de que hemos calculado una diferencial y todo eso y simplemente consideramos la *aproximación*, podemos pensar en ella como en una función definida "alrededor de (a, b)"; llamando $h_1 = x - a$ y $h_2 = y - b$ (los incrementos

de la variable cerca del punto), queda:

$$z = f(a,b) + -2a(x-a) + -2b(y-b) = (1-a^2-b^2) - 2a(x-a) - 2b(x-b),$$

que es una *función lineal*: una ecuación de primer grado en (x, y, z), es decir… un *plano*. Si os fijáis, la expresión de arriba tiene la misma estructura que

$$y = f(a) + f'(a)(x-a), \tag{9}$$

lo único que, en lugar de una variable hay dos y en lugar de la derivada, se calculan las parciales. La expresión (9) es, recuérdese, la de la *recta tangente a una función en* a, lo que nos ha quedado en dimensión 2 (tres si se cuenta la z) es *el plano tangente a la gráfica de* $f(x, y)$ *en el punto* (a, b). De hecho, esto es general:

Teorema 34. *Si* $f : X \to \mathbb{R}$ *es diferenciable en un punto* $a \in X \subset \mathbb{R}^2$*, entonces la expresión*

$$z = f(a) + \left.\frac{\partial f}{\partial x}\right|_a x + \left.\frac{\partial f}{\partial y}\right|_a y$$

es la ecuación del plano tangente a la gráfica de f *sobre punto* a.

Siguiendo con el paraboloide, en el punto $(x, y) = (0, 0)$ calculamos en su día la diferencial, que es la aplicación nula (es decir, las dos parciales son 0 en $(0, 0)$). Por tanto, el plano tangente al paraboloide $z = -x^2 - y^2 + 1$ sobre $(0, 0)$ tiene por ecuación

$$z = 1$$

que es, claro, un plano *horizontal*: el punto $(0, 0, 1)$ es el máximo del paraboloide y ahí su gráfica es "plana".

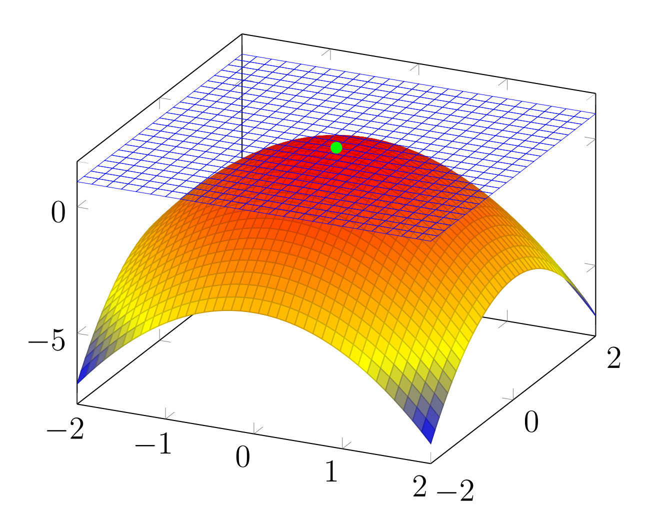


Figura 3. Paraboloide y plano tangente en $(0, 0, 1)$.

6. Curvas de nivel y vector gradiente

Otra manera de representar una función de 2 variables es, como en la Figura 2, utilizando las curvas de nivel: estas representan *líneas* en las que una función $f(x,y)$ toma el mismo valor. Son *exactamente* las curvas de nivel de un mapa: en estos, cada curva indica una altura constante, cuando se representa una función, se puede tomar cada valor de la función como una "altura" y *conectar* los puntos que están a la misma altura, de esta manera se genera un "mapa" en el plano, que representa, de otra manera, la gráfica de $f(x,y)$.

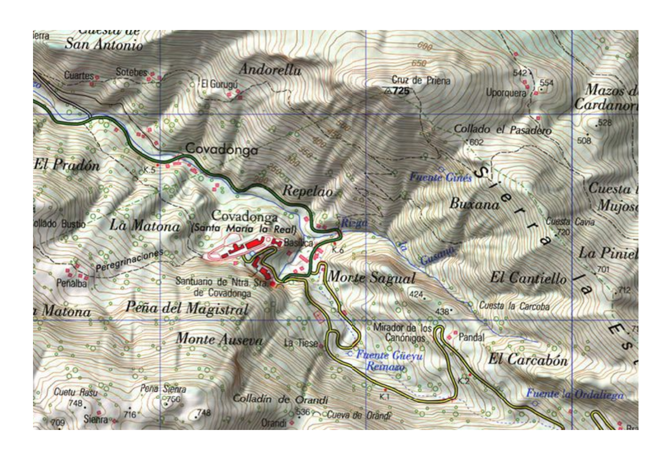


FIGURA 4. Curvas de nivel de un mapa cartográfico.

Las curvas de nivel, así, son las zonas (curvas) en las que $f(x,y)$ es constante. Es decir, la curva de nivel de "altura" c viene dada por la ecuación

$$f(x,y) = c.$$

Como se ve, esta función define una curva $\gamma(t)$ (se ve, no lo hemos demostrado, lo haremos más adelante). Supongamos que esta curva es diferenciable, etc.... Su ecuación será $\gamma(t) = (\gamma_1(t), \gamma_2(t))$. Si consideramos la función

$$h(t) = f(\gamma_1(t), \gamma_2(t)),$$

que a cada "instante" le asigna el valor de f en el punto correspondiente de la curva, es obvio que $h(t) = c$ cualquiera que sea t (pues la curva es exactamente el sitio en que f vale c). Es decir, $h(t)$ es una función constante y por tanto tiene derivada 0. ¿Cómo se calcula su derivada? Con la regla de la cadena, pues $h = f \circ \gamma$. Por tanto

$$h'(t) = D_{\gamma(t)} f \circ D_t(\gamma)(t)$$

pero γ es una función de $\mathbb{R}$ a $\mathbb{R}^2$, así que su diferencial tiene dos filas y una columna, y f es justo lo dual, así que tiene una fila y dos columnas. La jacobiana de γ es

$$J_t(\gamma) = \begin{pmatrix} \gamma_1{}'(t) \\ \gamma_2{}'(t) \end{pmatrix}$$

mientras que la de f es

$$J_{\gamma(t)}(f) = \begin{pmatrix} \frac{\partial f}{\partial x}\Big|_{\gamma(t)} & \frac{\partial f}{\partial y}\Big|_{\gamma(t)} \end{pmatrix}$$

de manera que

$$h'(t) = \begin{pmatrix} \frac{\partial f}{\partial x}\Big|_{\gamma(t)} & \frac{\partial f}{\partial y}\Big|_{\gamma(t)} \end{pmatrix} \begin{pmatrix} \gamma_1{}'(t) \\ \gamma_2{}'(t) \end{pmatrix} = 0$$

donde la igualdad a 0 es precisamente el hecho de que $h(t)$ es constante en la curva.

Acabamos de demostrar:

Teorema 35. *El vector de las derivadas parciales de una función de $\mathbb{R}^2$ en $\mathbb{R}$ es perpendicular a las curvas de nivel de esta función.*

De hecho, si uno se fija en el plano de Covadonga (Figura 4), especialmente en la zona nordeste, se observa que las curvas de nivel…son perpendiculares a la *pendiente máxima*: esto es especialmente visible en la zonas donde la pendiente es muy alta (donde la densidad de las curvas de nivel es muy grande): se ve claramente que las curvas de nivel son “muy paralelas” (además de estar muy juntas): de hecho, son paralelas porque porque la pendiente “cae” (o sube) en la dirección perpendicular.

Esto es general, en cualquier dimensión.

Definición 65. Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función derivable en el punto $a \in X$, interior de X. Se llama *vector gradiente* de f en a al vector “de las derivadas parciales”:

$$\nabla_a(f) = \begin{pmatrix} \frac{\partial f}{\partial x_1}\Big|_a & \dots & \frac{\partial f}{\partial x_n}\Big|_a \end{pmatrix}.$$

Se llama *campo gradiente* al vector gradiente “en todos los puntos” en que existe.

Definición 66. Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función de un abierto de $\mathbb{R}^n$ en $\mathbb{R}$. Se llaman *conjuntos de nivel* de f a los subconjuntos de la forma

$$\left\{ f(x_1, \dots, x_n) = c \middle| \, c \in \mathbb{R} \right\},$$

es decir, a los conjuntos en que f es constante.

Se tiene el resultado general siguiente:

Teorema 1. *Sea $f : \mathbb{X} \subset \mathbb{R}^n \to \mathbb{R}$ una función diferenciable en un abierto X y $a \in X$ un punto de X. El gradiente de f en a es perpendicular al (único) conjunto de nivel de f que pasa por a.*

Y, lo que es más: el gradiente *indica la dirección de máxima pendiente.*

TEOREMA 36. *Sea $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ una función del plano diferenciable en un punto a interior a X. Sea $D_a f$ su diferencial. Considérese la aplicación:*

$$p(t) = D_a f \begin{pmatrix} \cos(t) \\ \mathrm{sen}(t) \end{pmatrix}$$

definida en el intervalo compacto $[-\pi, \pi]$ (a cada vector de la circunferencia se le envía en el valor de la diferencial aplicada a ese vector). El máximo de p se da cuando el vector $(\cos(t), \mathrm{sen}(t))$ está en la misma dirección que $\nabla_a(f)$.

De hecho el enunciado es cierto en $\mathbb{R}^n$, pero es simplemente más pesado de enunciar (habría que hablar de variedades, y no quiero hacerlo). La función $p(t)$ de enunciado se podría llamar "pendiente en la dirección de ángulo t", por eso la he denotado con la letra p.

7. Derivadas de orden superior

Aunque solo las vamos a utilizar para los problemas de optimización, las *funciones derivadas parciales* pueden ser susceptibles de diferenciación y de dervación parcial (no dejan de ser funciones por el hecho de ser unas especiales). Considérese una función $f : X \subset \mathbb{R}^n \to \mathbb{R}$ de un abierto de $\mathbb{R}^n$ en $\mathbb{R}$,

$$f(x_1, \dots, x_n)$$

diferenciable. Sus derivadas parciales

$$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$$

se pueden considerar funciones también de X en $\mathbb{R}$,

$$\frac{\partial f}{\partial x_1}(a) := \left.\frac{\partial f}{\partial x_1}\right|_a, \dots, \frac{\partial f}{\partial x_n}(a) := \left.\frac{\partial f}{\partial x_n}\right|_a.$$

y pueden perfectamente ser diferenciables, y así sucesivamente... Se utiliza *más o menos* la siguiente nomenclatura:

DEFINICIÓN 67. Una función $f : X \subset \mathbb{R}^n \to \mathbb{R}$ de un abierto de $\mathbb{R}^n$ en $\mathbb{R}$ es *diferenciable k veces con continuidad* si es diferenciable, sus derivadas parciales lo son y son continuas y las de estas también, y son continuas y así hasta k veces, con continuidad.

Ejemplo: la derivada parcial de f con respecto a x_1:

$$\frac{\partial f}{\partial x_1}$$

podría ser diferenciable. Sus parciales:

$$\frac{\partial \frac{\partial f}{\partial x_1}}{\partial x_1}, \dots, \frac{\partial \frac{\partial f}{\partial x_1}}{\partial x_n}$$

no se escriben así. Como estamos "derivando" varias veces, lo indicamos poniendo el orden de derivación en el signo ∂ y como hemos de indicar con respecto a qué variables, lo decimos abajo:

$$\frac{\partial^2 f}{\partial x_1^2}, \dots, \frac{\partial^2 f}{\partial x_n \partial x_1},$$

nótese cómo cuando se deriva varias veces con respecto a la misma variable, se pone un exponente *en la variable.*

Por otro lado, resulta fundamental el siguiente teorema:

TEOREMA 37 (Regla de Schwartz). *Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función doblemente diferenciable con continuidad, y sean x_i y x_j dos coordenadas. Entonces:*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i},$$

es decir, en "buenas condiciones" (que se cumplirán en todos los ejemplos), las derivadas cruzadas coinciden.

Esta regla simplifica el cálculo de las derivadas parciales de orden superior (en cada orden hace falta calcular "la mitad" de derivadas parciales cruzadas, no todas ellas). La condición de continuidad es, como se dice en el enunciado, la continuidad es necesaria, pero salvo aparición de denominadores u otros problemas, se verificará "siempre". El (contra)ejemplo clásico (en el cual hay un denominador, claro) es:

$$f(x, y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} \text{ si } (x, y) \neq (0, 0) \\ 0 \text{ si } (x, y) = (0, 0) \end{cases}$$

que es derivable dos veces en el origen, *pero las parciales no son continuas en el origen.*

7.1. El Hessiano. Especial importancia tiene la matriz que se obtiene al calcular las derivadas parciales segundas de una función dos veces diferenciable con continuidad:

DEFINICIÓN 68. Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función dos veces diferenciable con continuidad en un punto $a \in X$. Se denomina *hessiano* de f en a a la matriz (simétrica) de sus derivadas parciales de segundo orden:

$$H_a(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

Como en condiciones normales (cuando las derivadas parciales segundas son continuas, cosa que ocurre *generalmente*) el hessiano, por la

Regla de Schwartz es una matriz simétrica, se puede entender que *define una forma bilineal simétrica* en el espacio vectorial $\mathbb{R}^n$. De hecho, uno siempre puede calcular el siguiente producto de vectores, dados $u, v \in \mathbb{R}^n$:

$$(10) \qquad \begin{pmatrix} u_1 & \dots & u_n \end{pmatrix} H_a(f) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = H_a(f)(u,v).$$

8. Función implícita, inversión de derivadas parciales, "Taylor" implícito

So, what about this?

9. Extremos locales ó "relativos"

Uf: dijimos hace mucho tiempo que el Cálculo está orientado, entre otras cosas, a los problemas de optimización. Pero no hemos optimizado más que un par de tonterías en el primer tema. ¿Qué hay de todo eso?

Consideremos una función $f : X \to \mathbb{R}$ donde X es un abierto de $\mathbb{R}^n$. Vamos a tratar de encontrar condiciones que determinen que un punto $a \in X$ sea un máximo o un mínimo de f, o al menos un máximo o mínimo "local".

9.1. Primera condición: diferencial nula. Observemos la Figura 5, que representa la función $f(x,y) = 1-x^2-y^2-2\,\mathrm{sen}(x+y)$ en el rectángulo $[-2,2]\times[-2,2]$. Está claro que hay un punto (x_0,y_0) (cuyas coordenadas desconocemos todavía) que es *especial*: cerca de él los valores de $f(x,y)$ son menores que $f(a,b)$ —o al menos menores o iguales—. Este punto es un *máximo local*. Si en lugar de $f(x,y) \leq f(a,b)$ se tiene

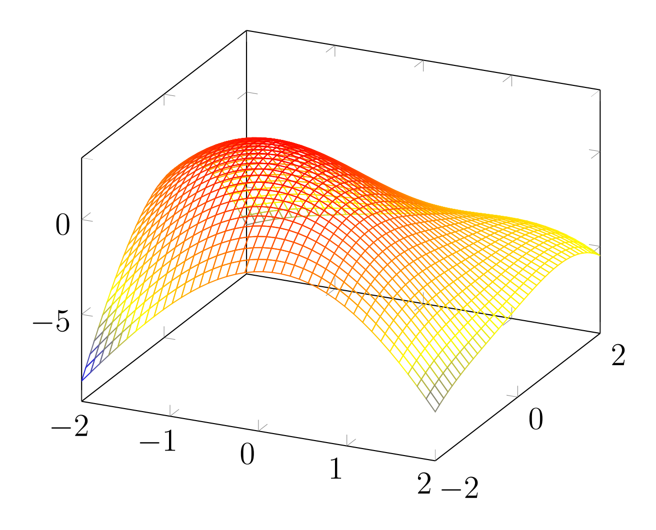


FIGURA 5. Un máximo *local*.

que $f(x,y) \geq f(a,b)$ *cerca* (es decir, *en un entorno*) de (a,b), entonces

se habla de (claro) un *mínimo local*. Estas dos nociones caen bajo el concepto de *extremo*:

DEFINICIÓN 69. Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función definida en un conjunto de $\mathbb{R}^n$ y $P \in X$ un punto de X. Se dice que P es un *máximo local* de f en X si existe una bola centrada en P, digamos $B_r(P)$ tal que $f(x) \leq f(P)$ para todo $x \in X \cap B_r(P)$. Se dice que es un *mínimo local* si $f(x) \geq f(P)$ para todo $x \in X \cap B_r(P)$. En cualquiera de los dos casos, se dice que P es un *extremo local* de f.

El hecho de que P sea un extremo local de f *no implica que P sea un extremo global de f*. No hay que insistir mucho en esto, basta observar la Figura 6: el origen (en este caso sí es el origen) es claramente un *máximo local*, pues *cerca* de $(0, 0)$ la función vale menos de lo que vale en $(0, 0)$, pero *obviamente* el origen no es un máximo global, pues como se ve, hay muchos puntos cerca del borde en los cuales el valor de f es mayor que $f(0, 0)$.

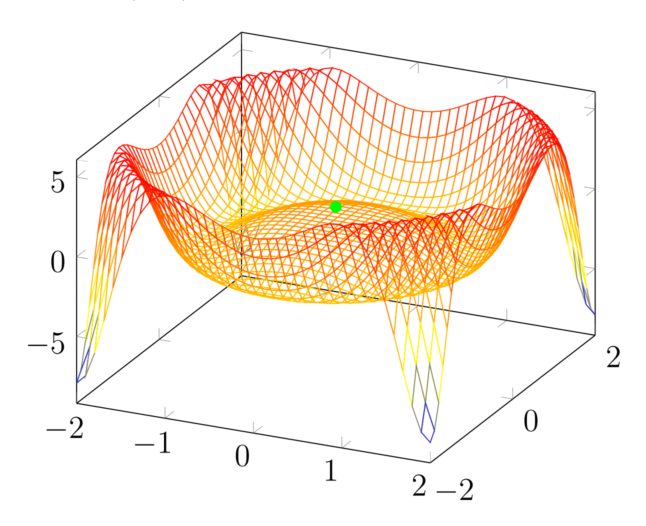


FIGURA 6. Otro máximo *local*.

DEFINICIÓN 70. Se dice que un punto $P \in X$ es un *máximo global* de $f : X \to \mathbb{R}$ si $f(P) \geq f(x)$ para todo $x \in X$. Se dice que es un *mínimo global* si la condición es $f(P) \leq f(x)$. En cualquier caso, se habla de *extremo* o de *extremo global*.

El problema es que el conjunto X puede ser muy raro. La búsqueda de máximos y mínimos *siempre comienza* por los extremos locales, pues son los primeros puntos susceptibles de serlo. Si no lo son, entonces se ha de recurrir a técnicas más sofisticadas.

En los ejemplos de las Figuras 5 y 6, se puede observar claramente que en el máximo, *el plano tangente a la gráfica de f es horizontal*. Esto *debería ser obvio*: si un punto P es un máximo *local* y la función f es diferenciable, entonces al dibujar el plano tangente a f por $(P, f(P))$,

este no puede estar inclinado, pues eso querría decir que en alguna dirección la gráfica de f crecería (y estamos suponiendo que P es un máximo local, la función, cerca de P, *no crece*). Del mismo modo, en un mínimo, el plano tangente tiene que ser horizontal. De hecho, esto es general:

TEOREMA 38. *Sea* $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ *una función de* 2 *variables y* $a \in X$ *un punto interior de* X*. Si* a *es un extremo local de* f*, entonces el plano tangente a* f *en* $(a, f(a))$ *es horizontal.*

DEMOSTRACIÓN. Supongamos que no. Sean $a = (x_0, y_0)$ las coordenadas de a. Entonces el plano tangente tendrá ecuación $z = A(x - x_0) + B(y - y_0) + f(a)$, lo cual quiere decir que existe un infinitésimo d tal que

$$f(a + h) = f(a) + Ah_1 + Bh_2 + \|h\|\mathrm{d}(h).$$

Supongamos que $A \neq 0$ (y pongamos que $A > 0$, si es $A < 0$ se cambia de signo todo). Cuando h es muy pequeño, el $\|h\|\mathrm{d}(h)$ es mucho más pequeño que Ah_1, así que, esencialmente,

$$f(a + (h_1, 0)) = f(a) + Ah_1,$$

pero esto quiere decir que $f(a+(h_1, 0)) > f(a)$ (pues A y h_1 se pueden tomar positivos) para puntos todo lo cerca que se quiera de a. Por tanto, A no puede ser distinto de 0. Lo mismo pasa con B. □

Pero no nos olvidemos de que *el plano tangente viene determinado por la diferencial, o si se quiere, por las parciales.* El hecho de que el plano tangente sea horizontal en $(a, f(a))$ no significa más que $D_a f$ es la aplicación nula, es decir $J_a(f) = (0\,0)$ —la matriz jacobiana es nula—, es decir:

TEOREMA 39. *Si* $(a, f(a))$ *es un extremo local de* $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ *en un punto interior de* X*, entonces*

$$\left.\frac{\partial f}{\partial x}\right|_a = 0, \quad \left.\frac{\partial f}{\partial y}\right|_a = 0.$$

De hecho, esto es general para todas las dimensiones:

TEOREMA 40. *Si* $(a, f(a))$ *es un extremo local de* $f : X \subset \mathbb{R}^n \to \mathbb{R}$ *en un punto interior de* X*, entonces*

$$\left.\frac{\partial f}{\partial x_1}\right|_a = 0, \left.\frac{\partial f}{\partial x_2}\right|_a = 0, \dots, \left.\frac{\partial f}{\partial x_2}\right|_a = 0.$$

Pero, *cuidado*, no basta esta condición para garantizar que un punto es un extremo local. En la Figura 7 se observa un ejemplo clásico (el denominado paraboloide hiperbólico, comúnmente conocido como *silla de montar*) de función que en $(0, 0)$ tiene parciales nulas pero para la que $(0, 0)$ no es ni máximo local ni mínimo local. Se ve que en una dirección la función crece y en otra decrece. La función es $f(x, y) =$

$x^2 - y^2$, que claramente "crece" en la dirección x y "decrece" en la dirección y.

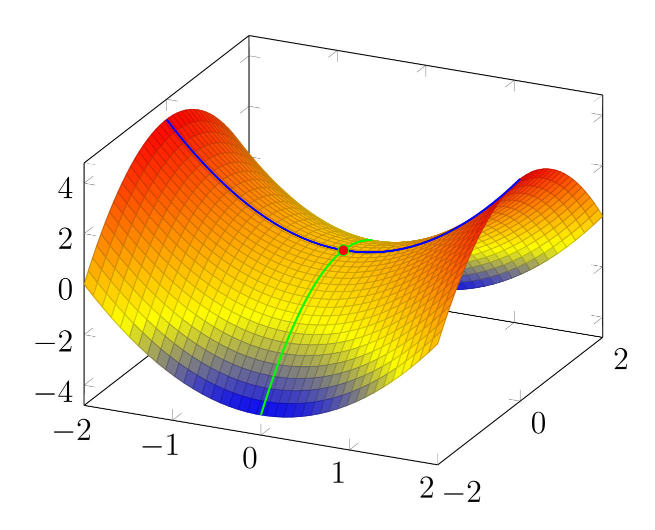


FIGURA 7. Punto de silla.

Una condición suficiente para garantizar un extremo local es la siguiente en $\mathbb{R}^n$:

TEOREMA 41. *Sea $H_a(f)$ el hessiano de f en un punto $a \in X \subset \mathbb{R}^n$ interior del dominio de definición de $f : X \to \mathbb{R}$. Si $H_a(f)$ define una forma bilineal simétrica definida positiva, entonces a es un* mínimo local. *Si $H_a(f)$ define una forma bilineal simétrica definida negativa, entonces a es un máximo local.*

Pero no siempre ocurre así: considérese $f(x, y) = (x - y)^2 = x^2 - 2xy + y^2$. Está claro (o debería estarlo) que $(0, 0)$ es un extremo local (pues $f(0, 0) \leq f(x, y)$ para cualquier (x, y)) y también global. Sin embargo,

$$H_{(0,0)}(f) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix},$$

que no es una forma definida, pues

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0,$$

mientras que si fuera definida, el producto de $(1, 1)$ consigo mismo usando la matriz cuadrada del hessiano debería ser distinto de 0.

DEFINICIÓN 71. Sea $f : X \subset \mathbb{R}^n \to \mathbb{R}$ una función y $a \in X$ un punto interior de X en el que f es diferenciable. Se dice que a es un *punto crítico* de f si $D_a(f) = 0$, es decir, si la diferencial es la aplicación nula. En otras palabras, si *todas* las derivadas parciales de f en a son nulas.

Los puntos críticos son "lo que uno busca" para encontrar extremos. A la hora de saber si es máximo o mínimo, lo primero que se hace es estudiar el Hessiano (si es que f es dos veces diferenciable con continuidad).

En la práctica, se utiliza uno de los siguientes resultados:

TEOREMA 42. *Si $a \in X$ es un extremo de $f : X \subset \mathbb{R}^n \to \mathbb{R}$ (se supone que a es un punto interior de X). Si f es dos veces diferenciable con continuidad y $H_a(f)$ es su hessiano en a, entonces:*

- *Si todos los autovalores de $H_a(f)$ son estrictamente positivos, entonces a es un mínimo local.*
- *Si todos los autovalores de $H_a(f)$ son estrictamente negativos, entonces a es un máximo local.*
- *En cualquier otro caso, no se sabe.*

A veces puede ser más sencillo no calcular los autovalores, sino utilizar los *menores principales* (es decir, los determinantes que se obtienen tomando las sucesivas matrices cuadradas que comienzan en el elemento superior izquierdo, hacia abajo y la derecha).

TEOREMA 43. *Como en el teorema anterior, sea $H_a(f)$ el hessiano de f en a (en las mismas condiciones). Sean $\Delta_1, \dots, \Delta_n$ los menores principales de $H_a(f)$. Entonces:*

- *Si $\Delta_1 > 0, \dots, \Delta_n > 0$ (todos son estrictamente positivos), entonces a es un máximo local.*
- *Si $\Delta_1 < 0, \Delta_2 > 0, \dots, (-1)^n\Delta_n > 0$ (es decir, los menores tienen signos alternos* empezando por negativo *y son no nulos), entonces a es un máximo.*
- *En cualquier otro caso, no se sabe.*

Téngase en cuenta que el Teorema 42 y el 43 son *equivalentes*: si uno decide, el otro también y si uno no lo hace, el otro tampoco (así que, por favor, no intentar "usar uno cuando el otro falla").

Y ¿qué pasa en los casos en que *no se sabe*? Pues que habrá que pensar un poco: habitualmente un razonamiento físico o un estudio más detallado de la función en cuestión puede concluir. En la Figura 8 se puede ver que el origen (marcado en verde) es un punto crítico (de hecho, es un mínimo local, aunque no sea mínimo estricto) pero la función $f(x, y) = (x + y)^2$ tiene por hessiano en $(0, 0)$ la matriz

$$H_{(0,0)}(f) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

cuyos menores principales son $\Delta_1 = 2, \Delta_2 = 0$, así que el criterio de los menores no decide. Pero en este casso concreto, uno "***sabe***[2]" que la función $f(x, y) = (x + y)^2$ *no puede ser negativa*, pues es un cuadrado, así que, si en un punto vale 0, entonces en ese punto la función tiene

[2]O al menos, *debería*

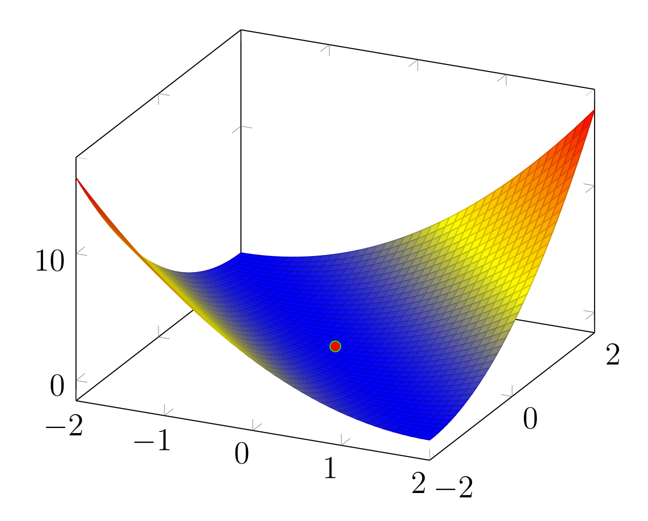


FIGURA 8. Punto crítico con hessiano no definido.

un mínimo (puede que no estricto, pero sí un mínimo). Por tanto, el punto $(0, 0)$ es un mínimo (de hecho, global) de f.

10. Optimización restringida: Lagrange

Una vez que uno sabe calcular extremos locales de funciones de varias variables, *aterriza en el mundo real* y se da cuenta de que, casi siempre, los problemas de optimización son de naturaleza distinta: se buscan puntos óptimos de funciones *sujetos a condiciones*. El ejemplo más sencillo quizás es:

EJEMPLO 17. Se tiene un alambre de 64cm. Calcular las dimensiones de un paralelogramo limitado por ese alambre que tenga área máxima.

El problema de este ejemplo es sencillo de trasladar a ecuaciones. Si un paralelogramo tiene dimensiones $l \times h$ (largo por ancho), entonces si P es el perímetro y S el área que encierra

$$P(l, h) = 2(l + h), \quad S(l, h, a) = lh$$

Y el problema concreto consiste en hallar el máximo de $S(l, h)$ *condicionado* a la restricción $P(l, h) = 64$ (asumimos que l y h están en centímetros). ¿Cómo proceder?

Es importante darse cuenta de que el problema *no tiene nada que ver con encontrar un extremo de S*: pues —evidentemente— el área solo tiene mínimos (ademas estos ocurren cuando una de las dimensiones es 0) y no tiene máximos. . .

Fijémonos en la Figura 9, que muestra los valores de una función (como puntos de color) $f(x, y)$ y la gráfica de *una curva de nivel* de otra función $g(x, y)$. En este caso, $f(x, y) = x + y$ y $g(x, y) = x^2 + y^2$ y en la curva que se dibuja, $x^2 + y^2 = 64$ (una circunferencia de radio

8). Supongamos que se trata de buscar el máximo de $f(x, y) = x + y$ *condicionado a que* $x^2 + y^2 = 64$, es decir, el máximo de f "sobre la curva".

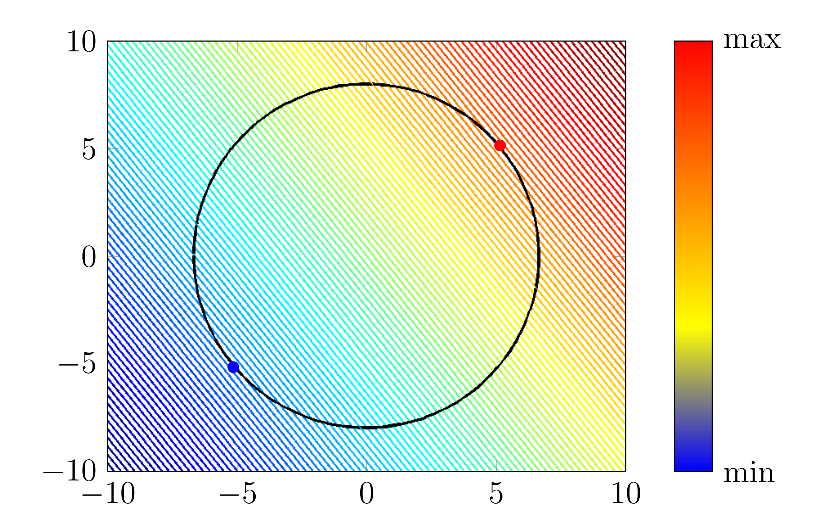


FIGURA 9. Optimización condicionada: el valor máximo de $f(x, y) = x + y$ en la circunferencia $x^2 + y^2 = 64$ es el punto "rojo", el mínimo el "azul": donde las rectas $x + y = c$ son tangentes a la circunferencia.

En la figura se han marcado los dos extremos, el máximo (en rojo) y el mínimo (en azul). Los colores indican los valores de la función $f(x, y) = x + y$ (es decir, sus curvas de nivel, que son las rectas $x + y = c$), hacia el azul son menores y hacia el rojo son mayores.

Si uno se fija en la Figura 10, que es la ampliación de una parte de la anterior, se observa que la curva $g(x, y) = 0$ (que es una *curva de nivel* de $g(x, y)$) es *transversal* a las curvas de nivel de $f(x, y) = c$ (que son las rectas de color). Como el campo gradiente de f es perpendicular a las curvas de nivel de f y el de g a las de esta, resulta que en este punto, *los gradientes van en direcciones distintas*. Es sencillo darse cuenta de que, como el gradiente ∇f indica la dirección de crecimiento de $f(x, y)$, si uno se desplaza *sobre la curva* $g(x, y) = 0$, los valores de $f(x, y)$ crecen en una dirección (en la de ∇f) y decrecen en la contraria.

Por el contrario, en la zona del máximo, en la Figura 11, se observa que los vectores ∇g y ∇f *tienen la misma dirección* y, en este caso concreto, si uno "se desplaza" sobre la curva $g(x, y) = 0$ cerca del punto estudiado, los valores de $f(x, y)$ *decrecen en los dos sentidos* (así que este punto es un máximo para f con la condición $g(x, y) = 0$).

En realidad, lo que se puede concluir es lo siguiente:

TEOREMA 44. *Sea $f(x, y)$ una función de $\mathbb{R}^2$ en $\mathbb{R}$ y $g(x, y) - c = 0$ una condición. Se supone que f y g son diferenciables con continuidad. Para que un punto $P = (x_0, y_0)$ sea un extremo de f condicionado*

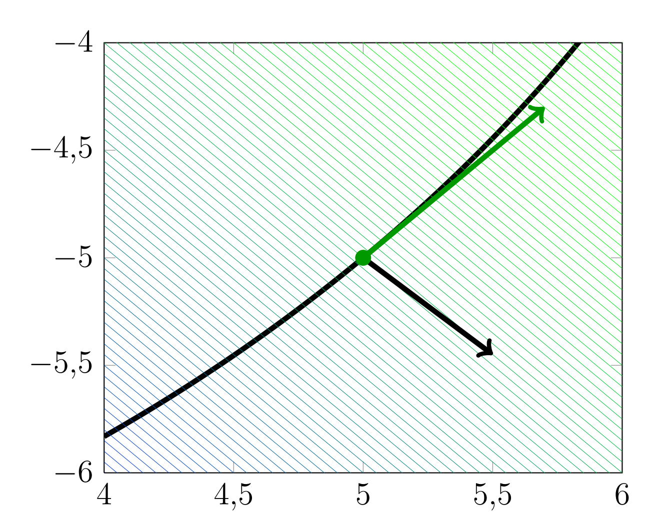


FIGURA 10. Zoom de una zona transversal entre $f(x, y) = c$ y $g(x, y) = 0$. La flecha verde es el gradiente de $f(x, y)$ en un punto P y la negra, el de $g(x, y)$ en el mismo punto. Tienen direcciones distintas. Cada línea de color es una línea de nivel de f, así que f no puede tener un extremo en P, pues los valores de f crecen en la dirección de la flecha.

a $g(x, y) - c = 0$ es necesario *que el gradiente de* f *y el de* g *sean proporcionales en* P.

De este resultado se deduce una manera de calcular los *posibles* extremos condicionados: se plantea la condición $g(x, y) - c = 0$ y las condiciones $\nabla f = \lambda \nabla g$, para cierto λ, que son:

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}, \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}$$

de manera que se tienen las ecuaciones:

$$\begin{aligned}
&\frac{\partial f}{\partial x}(x, y) = \lambda \frac{\partial g}{\partial x}(x, y) \\
&\frac{\partial f}{\partial y}(x, y) = \lambda \frac{\partial g}{\partial y}(x, y) \\
&g(x, y) - c = 0
\end{aligned}$$

que son tres ecuaciones con tres incógnitas: x, y y λ. Así que este sistema "se puede resolver" (pero por lo general, no se podrá resolver *a mano*).

Insisto en que lo que se obtiene cuando se resuelven esas ecuaciones son *posibles* extremos. Habrá que hacer en cada caso un análisis más fino para saber si los puntos obtenidos son máximos, mínimos o bien no son extremos, como en la figura

Para funciones de más variables, se tiene el mismo resultado:

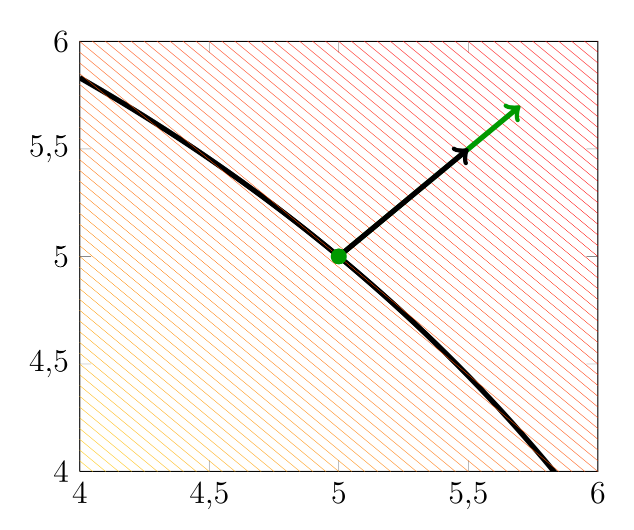


Figura 11. Zoom de una zona de tangencia entre $f(x, y) = c$ y $g(x, y) = 0$. La flecha verde es el gradiente de $f(x, y)$ en un punto P y la negra, el de $g(x, y)$ en el mismo punto. Tienen la misma dirección. Cerca de P, los valores de f en la curva "decrecen en los dos sentidos": si uno se mueve (sobre la curva) en una dirección o en otra, el valor de f decrece. El punto P es un máximo de f condicionado a $g(x, y) = 0$.

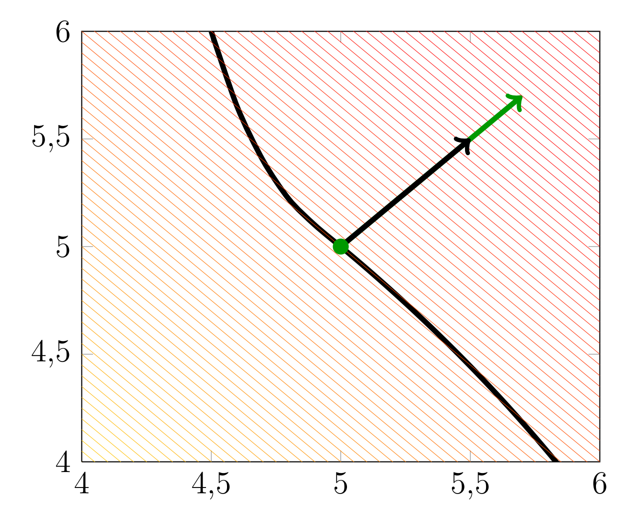


Figura 12. Un punto en que el gradiente de $f(x, y)$ y el de una curva $g(x, y)$ son paralelos pero que no es extremo condicionado (pues los valores de f crecen en ambas direcciones sobre la curva $g(x, y) = 0$.)

Teorema 45. *Sean $f, g : \mathbb{R}^n \to \mathbb{R}$ funciones de n variables diferenciables con continuidad. Se supone que $\nabla_{(x_1,\dots,x_n)} g$ es no nulo en todo el conjunto $g(x_1, \dots, x_n) = 0$. Si P es un extremo de f condicionado a*

$g(x_1,\ldots,x_n)=0$, *entonces existe* $\lambda\in\mathbb{R}$ *tal que*

$$\nabla_P f=\lambda\nabla_P g.$$

Es decir, para buscar los extremos condicionados con una sola condición, se ha de escribir

$$\left(\frac{\partial f}{\partial x_1}\bigg|_P,\ldots,\frac{\partial f}{\partial x_n}\bigg|_P\right)=\lambda\left(\frac{\partial g}{\partial x_1}\bigg|_P,\ldots,\frac{\partial g}{\partial x_n}\bigg|_P\right),$$

que, junto con $g(x_1,\ldots,x_n)=0$ da un sistema de $n+1$ ecuaciones con $n+1$ incógnitas.

Ejemplo 18. Se tiene una plancha de 100cm^2 de aluminio y se quiere hacer una caja paralelepipédica de volumen máximo. ¿Qué dimensiones tiene?

Llamemos a las variables, a,l,h (ancho, largo, alto). Se ha de optimizar la función $V(a,l,h)=alh$ con la restricción $S(a,l,h)=100$, donde la función S es

$$S(a,l,h)=2(al+ah+lh).$$

Calculemos los vectores gradientes:

$$\nabla_{(a,l,h)}V=(lh,ah,al),\ \nabla_{(a,l,h)}S=(2(l+h),2(a+h),2(a+l)).$$

Así que las ecuaciones quedan:

$$\begin{aligned}&lh=2\lambda(l+h)\\&ah=2\lambda(a+h)\\&al=2\lambda(a+l)\\&2(al+ah+lh)-100=0\end{aligned}$$

(obsérvese, y esto se hace siempre, que la constante de la restricción se pone en el primer miembro, por costumbre).

Utilizando WolframAlpha (téngase cuidado porque es un poco delicado con los nombres de las variables, conviene usar x,y,z,t o bien $x1,x2,\ldots$):

```
x*y=t*(x+y),x*z=t*(x+z),y*z=t*(y+z),2*(x*y+x*z+y*z)-100=0
```

(no hay ni siquiera que indicarle que queremos resolver el sistema), se obtienen dos resultados:

$$(a,l,h)=(-5\sqrt{\frac{2}{3}},-5\sqrt{\frac{2}{3}},-5\sqrt{\frac{2}{3}}),\ \ (a,l,h)=(5\sqrt{\frac{2}{3}},5\sqrt{\frac{2}{3}},5\sqrt{\frac{2}{3}})$$

y como las magnitudes no pueden ser negativas, solo es válida la segunda solución. ¿Por qué se sabe que es automáticamente un máximo? El volumen, por tanto, que se puede abarcar con un área de 100cm^2 es $125\cdot\frac{2\sqrt{2}}{3\sqrt{3}}\simeq 68{,}04cm^3\ldots$. Como debería saber todo el mundo, el paralelepípedo es un *cubo*.

Si en lugar de una condición se tienen varias, hay que utilizar varios parámetros, $\lambda_1,\ldots,\lambda_r$ (claro, uno por cada condición) e imponer la condición de que el gradiente de f sea una combinación lineal de los

gradientes de las g_i (esto es sencillo pero no trivial). Al final se tienen tantas ecuaciones como incógnitas.

TEOREMA 46. *Sean $f, g_1, \dots, g_m$ funciones de $\mathbb{R}^n$ en $\mathbb{R}$ diferenciables. Se supone que los vectores $\nabla_Q g_1, \dots, \nabla_Q g_m$ son linealmente independientes en el todos los puntos $Q = (x_1, \dots, x_n))$ del conjunto $g_1 = 0, \dots, g_m = 0$ (la* condición*). Si $P \in \mathbb{R}^n$ es un extremo de f condicionado a $g_1 = 0, \dots, g_m = 0$ entonces existen números reales $\lambda_1, \dots, \lambda_m$ tales que*

$$\nabla_P f = \lambda_1 \nabla_P g_1 + \cdots + \lambda_m \nabla_P g_m.$$

Así que, junto con las condiciones $g_1 = 0, \dots, g_m = 0$ se obtiene un sistema de $m + n$ ecuaciones con $m + n$ incógnitas, que "puede resolverse".

10.1. El Lagrangiano. . . En ocasiones se utiliza la siguiente nomenclatura. Se supone que $f(x_1, \dots, x_n)$ es una función de n variables que se quiere optimizar con las restricciones $g_1 = 0, \dots, g_m = 0$ (funciones de n variables también, las g_i). Todas las funciones se suponen diferenciables, etc. Entonces

DEFINICIÓN 72. El *lagrangiano* del problema de optimización de f con las restricciones $g_1 = 0, \dots, g_m = 0$ es la función de $n+m$ variables

$$\Lambda(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) =$$
$$= f(x_1, \dots, x_n) - \lambda_1 g_1(x_1, \dots, x_n) - \cdots - \lambda_m g_m(x_1, \dots, x_n)$$

Y la resolución del problema de optimizacón de f con las restricciones $g_i = 0$ se reduce a buscar los puntos críticos de Λ:

$$\nabla_P \Lambda(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = 0,$$

que da las ecuaciones de los Teoremas 45 y 46 cuando se escriben todas ellas (puesto que Λ es una función de $n + m$ variables, las x_i y las λ_j).

A las variables λ_j (o a sus valores una vez resuelto el problema de optimización) se las denomina *multiplicadores de Lagrange.*

10.2. Significado de los multiplicadores.

10.3. Optimización con desigualdades. En "la vida real" (si es que esto existe), la optimización no siempre está sujeta a condiciones de igualdad*. Por ejemplo

Fixme: *¿Dejar esto para que lo presenten? (activa/inactiva. . .)*

EJEMPLO 19. Se tiene una plancha de 100cm^2 de aluminio y se quiere hacer una caja paralelepipédica de volumen máximo de manera que el perímetro sea menor de 15cm ¿Qué dimensiones tiene?